

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
18 August 2005 (18.08.2005)

PCT

(10) International Publication Number
WO 2005/076010 A2

(51) International Patent Classification⁷: **G01N 33/68**

(21) International Application Number:
PCT/IN2005/000037

(22) International Filing Date: 7 February 2005 (07.02.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
173/DEL/2004 6 February 2004 (06.02.2004) IN
60/589,227 20 July 2004 (20.07.2004) US

(71) Applicant (for all designated States except US): **COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH** [IN/IN]; Rafi Marg, 110 001 New Delhi, Maharashtra (IN).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SACHDEVA, Gaurav** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **KUMAR, Kaushal** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **JAIN, Preti** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **BRAHMACHARI, Samir, Kumar** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, New Delhi 110 007 (IN). **RAMACHANDRAN, Srinivasan** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN).

(74) Agents: **BHOLA, Ravi** et al.; K & S Partners, 84-C, C6 Lane, Off Central Avenue, Sainik Farms, New Delhi 110 067 (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

Published:

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE PROTEINS OF THERAPEUTIC POTENTIAL

(57) Abstract: A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of a neural network software wherein the attributes are (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 ; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins.



WO 2005/076010 A2

COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE PROTEINS OF THERAPEUTIC POTENTIAL

Field of the present Invention

5 A computational method for identifying adhesin and adhesin-like proteins; computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

Background and prior art of the present invention

10 The progress in genome sequencing projects has generated a large number of inferred protein sequences from different organisms. It is expected that the availability of the information on the complete set of proteins from infectious human pathogens will enable us to develop novel molecular approaches to combat them. A necessary step in the successful colonization and subsequent manifestation of disease by microbial pathogens is the ability to adhere to host cells.

15 Microbial pathogens encode several proteins known as adhesins that mediate their adherence to host cell surface receptors, membranes, or extracellular matrix for successful colonization. Investigations in this primary event of host-pathogen interaction over the past decades have revealed a wide array of adhesins in a variety of pathogenic microbes. Presently, substantial information on the biogenesis of adhesins and the regulation of adhesin factors is available. One of the best understood mechanisms of bacterial adherence is attachment mediated by pili or fimbriae. Several afimbrial adhesins also have been reported. In addition, limited knowledge on the target host receptors also has been gained (Finlay, B.B. and Falkow, S 1997).

25 New approaches to vaccine development focus on targeting adhesins to abrogate the colonization process (Wizemann, *et al* 1999). However, the specific role of particular adhesins has been difficult to elucidate. Thus, prediction of adhesins or adhesin-like proteins and their functional characterization is likely to aid not only in deciphering the molecular mechanisms of host pathogen interaction but also in developing new vaccine formulations, which can be tested in suitable experimental model systems.

30 One of the best understood mechanisms of bacterial adherence is attachment mediated by pili or fimbriae. For example, FimH and PapG adhesins of *Escherichia coli* (Maurer, L., Orndorff, P.(1987), Bock, K., *et al.*(1985). Other examples of pili group adhesins include type IV pili in *Pseudomonas aeruginosa*, *Neisseria* species, *Moraxella* species, Enteropathogenic *Escherichia coli* and *Vibrio cholerae* (Sperandio V *et al* (1996).

Several afimbrial adhesins are HMW proteins of *Haemophilus influenzae* (van Schilfgaarde 2000), the filamentous hemagglutinin, pertactin, of *Bordetella pertussis* (Bassinet *et al* 2000), the BabA of *H. pylori* (Yu J *et al* 2002) and the YadA adhesin of *Yersinia enterocolitica* (Neubauer *et al* 2000). The intimin receptor protein (Tir) of Enteropathogenic *E. coli* (EPEC) is another type of adhesin (Ide T *et al* 2003). Other class of adhesins includes MrkD protein of *Klebsella pneumoniae*, Hia of *H. influenzae* (St Geme *et al* 2000), Ag I/II of *Streptococcus mutans* and SspA, SspB of *Streptococcus gordonii* (Egland *et al* 2001), FnbA, FnbB of *Staphylococcus aureus* and SfbI, protein F of *Streptococcus pyogenes*, the PsaA of *Streptococcus pneumoniae* (De *et al* 2003).

A known example of adhesins approved as vaccine is the acellular pertussis vaccine containing FHA and pertactin against *B. pertussis* the causative agent of whooping cough (Halperin, S *et al* 2003). Immunization with FimH is being evaluated for protective immunity against pathogenic *E. coli* (Langermann S *et al* 2000), in *Streptococcus pneumoniae*, PsaA is being investigated as a potential vaccine candidate against pneumococcal disease (Rapola, S *et al* 2003). Immunization results with BabA adhesin showed promise for developing a vaccine against *H. pylori* (Prinz, C *et al* 2003). A synthetic peptide sequence anti-adhesin vaccine is being evaluated for protection against *Pseudomonas aeruginosa* infections.

Screening for adhesin and adhesin like proteins by conventional experimental method is laborious, time consuming and expensive. As an alternative, homology search is used to facilitate the identification of adhesins. Although, this procedure is useful in the analysis of genome organization (Wolf *et al* 2001) and of metabolic pathways (Peregrin-Alvarez *et al* 2003, Rison *et al* 2002), it is somewhat limited in allowing functional predictions when the homologues are not functionally characterized or the sequence divergence is high. Assignment of functional roles to proteins based on this technique has been possible for only about 60% of the predicted protein sequences (Fraser *et al* 2000). Thus, we explored the possibility of developing a non-homology method based on sequence composition properties combined with the power of the Artificial Neural Networks to identify adhesins and adhesin-like proteins in species belonging to wide phylogenetic spectrum.

Twenty years ago, Nishikawa *et al* carried out some of the early attempts to classify proteins into different groups based on compositional analysis (Nishikawa *et al* 1983).

More recently, the software PropSearch was developed for analyzing protein sequences where conventional alignment tools fail to identify significantly similar sequences (Hobohm, U. and Sander, C 1995). PropSearch uses 144 compositional properties of protein sequences to detect possible structural or functional relationships between a new sequence and sequences in the database. Recently the compositional attributes of proteins have been used to develop softwares for predicting secretory proteins in bacteria and apicoplast targeted proteins in *Plasmodium falciparum* by training Artificial Neural Networks (Zuegge *et al* 2001).

Zuegge *et al* have used the 20 amino acid compositional properties. Their objective was to extract features of apicoplast targeted proteins in *Plasmodium falciparum*. This is distinct from our software SPAAN that focuses on adhesins and adhesin-like proteins involved in host-pathogen interaction.

Hobohm and Sander have used 144 compositional properties including isoelectric point and amino acid and dipeptide composition to generate hypotheses on putative functional role of proteins that are refractory to analysis using other sequence alignment based approaches like BLAST and FASTA. Hobohm and Sander do not specifically address the issue of adhesins and adhesin-like proteins, which is the focus of SPAAN

Nishikawa *et al* had originally attempted to classify proteins into various functional groups. This was a curiosity driven exercise but eventually lead to the development of a software to discriminate extra-cellular proteins from intracellular proteins. This work did not address the issue of adhesins and adhesin-like proteins, which is the focus of SPAAN.

Thus, none of the aforementioned research groups have been able to envisage the methodology of the instant application. The inventive method of this application provides novel proteins and corresponding gene sequences.

Adhesins and adhesin-like proteins mediate host-pathogen interactions. This is the first step in colonization of a host by microbial pathogens. Attempts Worldwide are focused on designing vaccine formulations comprising adhesin proteins derived from pathogens. When immunized, host will have its immune system primed against adhesins for that pathogen. When a pathogen is actually encountered, the surveillance mechanism will recognize these adhesins, bind them through antigen-antibody interactions and neutralize the pathogen through complement mediate cascade and other related clearance mechanisms. This strategy has been successfully employed in

the case of Whooping cough and is being actively pursued in the case of Pneumonia, Gastric Ulcer and Urinary tract infections.

Objects of the present Invention

The main object of the present invention is to provide a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential.

Another object of invention is to provide a method for screening the proteins with unique compositional characteristics as putative adhesins in different pathogens.

Yet, another object of the invention is providing the use of gene sequences encoding the putative adhesin proteins useful as preventive therapeutics.

Summary of the present Invention

A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of protein sequences using five attribute modules of software SPAAN, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 ; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

Detailed description of the present Invention

Accordingly, the present invention relates to a computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of protein sequences using five attribute modules of software SPAAN, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 ; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

In an embodiment of the present invention, wherein the invention relates to a computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of:

- a. computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, wherein the attributes

are software, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition,

- b. training the artificial neural Network (ANN) for each of the computed five attributes, and
- c. identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 .

In another embodiment of the present invention, wherein the invention relates to a method wherein the protein sequences is obtained from pathogens, eukaryotes, and multicellular organisms.

In an embodiment of the present invention, wherein the invention relates to a method, wherein the protein sequences are obtained from the pathogens selected from a group of organisms comprising *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum* and Severe Acute Respiratory Syndrome associated human coronavirus (SARS).

In yet another embodiment of the present invention, wherein the method of the invention is a non-homology method.

In still another embodiment of the present invention, wherein the invention relates to the method using 105 compositional properties of the sequences.

In still another embodiment of the present invention, wherein the invention relates to a method showing sensitivity of at least 90%.

In still another embodiment of the present invention, wherein the invention relates to the method showing specificity of 100%.

In still another embodiment of the present invention, wherein the invention relates to a method identifying adhesins from distantly related organisms.

In still another embodiment of the present invention, wherein the invention relates to the neural network has multi-layer feed forward topology, consisting of an input layer, one hidden layer, and an output layer.

In still another embodiment of the present invention, wherein the invention relates to the number of neurons in the input layer are equal to the number of input data points for each attribute.

In still another embodiment of the present invention, wherein the invention relates to the "P_{ad}" is a weighted linear sum of the probabilities from five computed attributes.

In still another embodiment of the present invention, wherein the invention relates to each trained network assigns a probability value of being an adhesin for the protein sequence.

5 In still another embodiment of the present invention, wherein the invention relates to a computer system for performing the method of claim 1, said system comprising a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on
10 which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

In still another embodiment of the present invention, wherein the invention relates to a set of 274 annotated genes encoding adhesin and adhesin-like proteins, having SEQ ID
15 Nos. 385 to 658.

In still another embodiment of the present invention, wherein the invention relates to a set of 105 hypothetical genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 659 to 763.

In still another embodiment of the present invention, wherein the invention relates to a set of 279 annotated adhesin and adhesin-like proteins of SEQ ID Nos. 1 to 279.

In still another embodiment of the present invention, wherein the invention relates to a set of 105 hypothetical adhesin and adhesin-like proteins of SEQ ID Nos. 280 to 384.

One more embodiment of the present invention, wherein the invention also relates to a fully connected multilayer feed forward Artificial Neural Network based on the
25 computational method as claimed in claim 1, comprising of an input layer, a hidden layer and an output layer which are connected in the said sequence, wherein each neuron is a binary digit number and is connected to each neuron of the subsequent layer for identifying adhesin or adhesin like proteins, wherein the program steps comprise:-
[a] feeding a protein sequence in FASTA format; [b] processing the sequence
30 obtained in step [a] through the 5 modules named A, C, D, H and M, wherein attribute A represents an amino acid composition, attribute C represents a charge composition, attribute D represents a dipeptide composition of the 20 dipeptides [NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI and HR], attribute H

represents a hydrophobic composition and attribute M represents amino acid frequencies in multiplets to quantify 5 types of compositional attributes of the said protein sequence to obtain numerical input vectors respectively for each of the said attributes wherein the sum of numerical input vectors is 105; [c] processing of the numerical input vectors obtained in step [b] by the input neuron layer to obtain signals, wherein the number of neurons is equal to the number of numerical input vectors for each attribute; [d] processing of signals obtained from step [c] by the hidden layer to obtain synaptic weighted signals, wherein the optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually; [e] delivering synaptic weighted signals obtained from step [d] to the output layer for assigning of a probability value for each protein sequence fed in step [a] as being an adhesin by each network module; [f] using the individual probabilities obtained from step [e] for computing the final probability of a protein sequence being an adhesin denoted by the P_{ad} value, which is a weighted average of the individual probabilities obtained from step [e] and the associated fraction of correlation which is a measure of the strength of the prediction.

In still another embodiment of the present invention, wherein the input neuron layer consists of a total of 105 neurons corresponding to 105 compositional properties.

In still another embodiment of the present invention, wherein the hidden layer comprises of neurons represented as 30 for amino acid frequencies, 28 for multiplet frequencies, 28 for dipeptide frequencies, 30 for charge composition and 30 for hydrophobic composition.

In still another embodiment of the present invention, wherein the output layer comprises of neurons to deliver the output values as probability value for each protein sequence.

Identification of novel adhesins and their characterization are important for studying host-pathogen interactions and testing new vaccine formulations. We have employed Artificial Neural Networks to develop an algorithm SPAAN (Software for Prediction of Adhesin and Adhesin-like proteins using Neural Networks) that can identify adhesin proteins using 105 compositional properties of a protein sequence. SPAAN could correctly predict well characterized adhesins from several bacterial species and strains. SPAAN showed 89% sensitivity and 100% specificity in a test data set that did not

contain proteins in the training set. Putative adhesins identified by the software can serve as potential preventive therapeutics.

The present invention provides a novel computational method for identifying adhesin and adhesin-like proteins of therapeutic potential. More particularly, the present invention relates to candidate genes for these adhesins. The invention further provides new leads for development of candidate genes, and their encoded proteins in their functional relevance to preventive approaches. This computational method involves calculation of several sequence attributes and their subsequent analyses lead to the identification of adhesin proteins in different pathogens. Thus, the present invention is useful for identification of the adhesin proteins in pathogenic organisms. The adhesin proteins from different genomes constitute a set of candidates for functional characterization through targeted gene disruption, microarrays and proteomics. Further, these proteins constitute a set of candidates for further testing in development of preventive therapeutics. Also, are provided the genes encoding the candidate adhesin proteins.

The present method offers novelty in the principles used and the power of Neural Networks to identify new adhesins compared to laborious and time consuming conventional methods. The present method is based on compositional properties of proteins instead of sequence alignments. Therefore this method has the ability to identify adhesin and adhesin like proteins from bacteria belonging to a wide phylogenetic spectrum. The predictions made from this method are readily verifiable through independent analysis and experimentation. The invention has the potential to accelerate the development of new preventive therapeutics, which currently requires high investment in terms of requirement of skilled labor and valuable time.

The present invention relates to a computational method for the identification of candidate adhesin proteins of therapeutic potential. The invention particularly describes a novel method to identify adhesin proteins in different genomes of pathogens. These adhesin proteins can be used for developing preventive therapeutics.

Accordingly, a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential which comprises calculation of 105 compositional properties under the five sequence attributes, namely, Amino Acid frequency, Multiplet frequency, dipeptide frequency, charge composition and hydrophobic composition; and then training Artificial Neural Network (ANN, Feed Forward Error Back Propagation)

using these properties for differentiating between adhesin and non-adhesin class of proteins. This computational method involves quantifying 105 compositional attributes of query proteins and qualifying them as adhesins or non-adhesins by a P_{ad} value (Probability of being an adhesin). The present invention is useful for identification of adhesin and adhesin-like proteins in pathogenic organisms. These newly identified adhesin and adhesin-like proteins constitute a set of candidates for development of new preventive therapeutics that can be tested in suitable experimental model systems readily. In addition, the genes encoding the candidate adhesin and adhesin-like proteins are provided.

The invention provides a set of candidate adhesin and adhesin-like proteins and their coding genes for further evaluation as preventive therapeutics. The method of invention is based on the analysis of protein sequence attributes instead of sequence patterns classified to functional domains. Present method is less dependent on sequence relationships and therefore offers the potential power of identifying adhesins from distantly related organisms. The invention provides a computational method, which involves prediction of adhesin and adhesin-like proteins using Artificial Neural Networks. The proteins termed adhesin were found to be predicted with a high probability ($P_{ad} = 0.51$) in various pathogens. Some adhesin sequences turned out to be identical or homologous to proteins that are antigenic or implicated in virulence. By this approach, proteins could be identified and short-listed for further testing in development of new vaccine formulations to eliminate diseases caused by various pathogenic organisms.

DESCRIPTION OF TABLES

Table 1: Output file format given by SPAAN.

Table 2: Organism Name, Accession number, Number of base pairs, Date of release and Total number of proteins.

Table 3. Prediction of well characterized adhesins from various bacterial pathogens using SPAAN.

Table 4. Analysis of predictions made by SPAAN on genome scans of a few selected pathogenic organisms.

Table 5: GI numbers and Gene IDs of new putative adhesins predicted by SPAAN in the genomes listed in Table 2.

Table 6: GI numbers and Gene IDs of hypothetical proteins predicted as putative adhesins by SPAAN in the genomes listed in Table 2.

Table 7: The list of 198 adhesins found in bacteria

Brief description of the accompanying drawings

5 **Figure 1** shows the Neural Network architecture

Figure 2 shows assessment of SPAAN using defined test dataset.

Figure 3 (a) shows Histogram plots of the number of proteins in the various P_{ad} value ranges are shown. **(b)** Pairwise sequence relationships among the adhesins were determined using CLUSTAL W and plotted on X-axis. Higher scores indicate similar pairs. **(c)** plot for non-adhesins. Data are plotted in the 4 quadrant format for clear inspection.

Software program was written in C Language and operated on Red Hat Linux 8.0 operating system. The computer program accepts input protein sequences in FastA format and produces a tabulated output. The output Table contains one row for each protein listing the probability outputs of each of the five modules, a weighted average probability of these five modules (P_{ad}), and the function of the protein as described in the input sequence file. This software is called SPAAN (A Software for Prediction of Adhesins and Adhesin-like proteins using Neural Networks) and a software copyright has been filed. Although this software has multiple modules, the running of these modules have been integrated and automated. The user only needs to run one command.

AAcompo.c:

Input: File containing protein sequences in the fasta format.

Output: File containing frequencies of all 20 AAs for each protein in one row.

25 **charge.c:**

Input: File containing protein sequences in the fasta format.

Output: File containing frequency of charged amino acids (R, K, E and D) and moments (up to 18th order) of the positions of charged amino acids.

hdr.c:

30 **Input:** File containing protein sequences in the fasta format.

Output: File containing frequencies of 5 groups of amino acids formed on the basis their Hydrophobicity and moments of their positions up to 5th order.

multiplets.c:

Input: File containing protein sequences in the fasta format.

5 **Output:** File containing fractions of multiplets of each of the 20 amino acids.

query-di pep.c:

Input: File.1 containing protein sequences in the fasta format.

File.2 containing list of the significant dipeptides in dipeptide analysis.

10 **Output:** File containing frequencies of the dipeptides listed in the input File.2 for each protein in the input File.1.

train.c:

Input: File containing following specifications –

1. Number of input and output parameters.
2. Number of nodes in the hidden layers.
- 15 3. Names of the training, validate and test data files.
4. Learning rate, coefficient of moment.
5. Maximum number of cycles for training.

Output: Outputs are as follows.

1. Output of the trained NN for the test data set.
- 20 2. Values of the weight connections in the trained NN.
3. Some extra information about training.

recognize.c:

Input: File containing following specifications –

1. Number of input and output parameters.
- 25 2. Number of nodes in the hidden layers.
3. Names of the query input file.
4. Name of the file containing values of the weight connections for trained NN.
5. Name of the output file.

30 **Output:** Outputs for the query entries calculated by the trained NN.

standard.c:

Input: File containing protein sequences in fasta format.

Output: File containing protein sequences in fasta format with all the new line characters removed lying within a sequence.

5 **filter.c:**

Input: File containing protein sequences in fasta format.

Output: File containing protein sequences from the input except those which are short in length (<50 AAs) and which contain any amino acid other than the 20 known amino acids.

10 ***The five attributes:***Amino Acid frequencies

Amino acid frequency $f_i = (\text{counts of } i^{\text{th}} \text{ amino acid in the sequence}) / l$; $i, = 1 \dots 20$, l is the length of the protein.

Multiplet frequency

15 Multiplets are defined as homopolymeric stretches $(X)_n$ where X is any of the 20 amino acids and n is an integer > 2 . After identifying all the multiplets, the frequencies of the amino acids in the multiplets were computed as

$$f_i(m) = (\text{counts of } i^{\text{th}} \text{ amino acid occurring as multiplet}) / l$$

Dipeptide frequencies

20 The frequency of a dipeptide (i, j) $f_{ij} = (\text{counts of } ij^{\text{th}} \text{ dipeptide}) / (\text{total dipeptide counts})$; i, j ranges from 1 to 20.

It has been found that dipeptide repeats in proteins are important for functional expression of the clumping factor present on *Staphylococcus aureus* cell surface that binds to fibrinogen (Hartford *et al* 1999). Thus we included the dipeptide frequency module. The total number of dipeptides is 400. For optimal training of Neural Network, the ratio of total number of input vectors to the total number of weight connections must be around 2 to avoid over fitting (Andrea *et al*). Therefore, we identified the dipeptides whose frequencies in the adhesin data set (469 proteins, see database construction) were significantly different from that in the non-adhesin dataset (703 proteins) using *t-test*. The frequencies of top 20 dipeptides (when arranged in the descending order of the p-values of *t-test*), were fed to the Neural Network. These dipeptides were (using single letter IUPAC-IUB code) NG, RE, TN, NT, GT, TT, DE,

25

30

ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI, AND HR. With frequency inputs for 20 dipeptides and 28 neurons in the 2nd layer, the total number of weight connections is 588, and is in keeping with the criterion of avoiding over fitting.

Charge composition

- 5 The input frequency of charged amino acids (R, K, E and D considering the ionization properties of the side chains at pH 7.2) given by $f_c = (\text{counts of charged amino acids}) / l$. Further, information on the characteristics of the distribution of the charged amino acids in a given protein sequence was provided by computing the moments of the positions of the occurrences of the charged amino acids. Since moments characterize
- 10 the patterns of distribution such as skewness and kurtosis (sharpness of the peak) we have used them to represent the distribution patterns of the charged residues in the sequence.

The general expression to compute moments of a given order; say 'i' is

$M_r = r^{\text{th}}$ order moment of the positions of charged amino acids

$$15 \quad = \sum \frac{(X_i - X_m)^r}{N}$$

Where, X_m = mean of all positions of charged amino acids

X_i = position of i^{th} charged amino acid

N = number of charged amino acids in the sequence

- The moments 2^{nd} to 19^{th} order were used to train the ANN constituting a total 20 inputs
- 20 in addition to frequency of charged amino acids and the length of the protein. The upper limit of 19^{th} order was set based on assessments of sensitivity and specificity on a small dataset of adhesins and non-adhesins. Moments of order greater than 19 were not useful in improvement of performance.

Hydrophobic composition

- 25 A given protein sequence was digitally transformed using the hydrophobic scores of the amino acids according to Brendel *et al.* (43). The scores for five groups of amino acids: (-8 for K, E, D, R), (-4 for S, T, N, Q), (-2 for P, H), (+1 for A, G, Y, C, W), (+2 for L, V, I, F, M).

Following inputs were given for each of the group

- 30 (a) $f_i = (\text{counts of } i^{\text{th}} \text{ group}) / (\text{total counts in the protein})$; i ranges from 1 to 5
- (b) $m_{ji} = j^{\text{th}}$ order moment of positions of amino acids in i^{th} group; j ranges from 2 to 5.

A total of 25 inputs representing the hydrophobic composition of a protein were fed to the Neural Network. The rationale for using moments was same as described in the section on charge composition inputs.

Taken together a total of 105 compositional properties of a given protein sequence were used to predict their adhesin characteristics.

The software PropSearch uses 144 compositional properties of protein sequences to detect possible structural or functional relationships between a new sequence and sequences in the database (Hobohm and Sander 1995). The approach defines protein sequence dissimilarity (or distance) as a weighted sum of differences of compositional properties such as singlet and doublet amino acid composition, molecular weight, isoelectric point (protein property search or PropSearch). Compositional properties of proteins have also been used for predicting secretory proteins in bacteria and apicoplast targeted proteins in *Plasmodium falciparum* (Zuegge, et al. 2001). The properties used here are statistical methods, principal component analysis, self-organizing maps, and supervised neural networks. In SPAAN, we have used 105 compositional properties in the five modules viz. Amino Acid frequencies, Multiplet frequencies, Dipeptide frequencies, Charge composition, Hydrophobic composition. The total of 105 properties used in SPAAN are 20 for Amino acid frequencies, 20 for Multiplets frequencies, 20 for Dipeptide frequencies (Top 20 significant dipeptides are used, based on *t-test*), 20 for Charge composition (frequency of charged amino acids (R, K, E and D) and moments of 2nd to 19th order), and 25 for Hydrophobic composition (Amino acids were classified into five groups (-8 for K, E, D, R), (-4 for S, T, N, Q), (-2 for P, H), (+1 for A, G, Y, C, W), (+2 for L, V, I, F, M). A total of 25 inputs consisted of the following: Frequency of each group, Moments of positions of amino acids in each group from 2nd to 5th order.

Neural Network

A feed forward error back propagation Neural Network was used. The program is a kind gift from Charles W. Anderson, Department of Computer Science, Colorado State University, Fort Collins, CO 80523, anderson@cs.colostate.edu

Neural Network architecture

The Neural Network used here has a multi-layer feed-forward topology. It consists of an input layer, one hidden layer and an output layer. This is a 'fully-connected' Neural Network where each neuron *i* is connected to each unit *j* of the next layer (Figure 1).

The weight of each connection is denoted by w_{ij} . The state I_i of each neuron in the input layer is assigned directly from the input data, whereas the states of hidden layer neurons are computed by the sigmoid function,

$$h_j = 1 / (1 + \exp -(w_{j0} + w_{ij} I_i)),$$

5 where, w_{j0} is the bias weight

The back propagation algorithm was used to minimize the differences between the computed output and the desired output. Ten thousand cycles (epochs) of iterations are performed. Subsequently, the best epoch with minimum error was identified. At this point the network produces approximate target values for a given input in the training
10 set.

A network was trained optimally for each attribute. Thus five networks were prepared. The schematic diagram (Figure 1) shows the procedure adopted. The number of neurons in the input layer was equal to the number of input data points for each attribute (for example 20 neurons for 20 numerical input vectors of the amino acid
15 composition attribute). The optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually. An upper limit for the total number of weight connections was set to half of the total number of input vectors to avoid over fitting as suggested previously (Andrea *et al*).

20 Computer programs to compute individual compositional attributes were written in C and executed on a PC under Red Hat Linux ver 7.3 or 8.0. The network was trained on the training set, checks error and optimizes using the validate set through back propagation. The validate set was different from the training set. Since, the number of well annotated adhesins were not many, we used the 'validate set' itself as test set for
25 preliminary evaluation of the performance and to obtain the fraction of correlation to compute the weighted average probability (P_{ad} value) described in the next section. The training set had 367 adhesins and 580 non-adhesins. The validate set had 102 adhesins and 123 non-adhesins. The adhesins were qualified with a digit '1' and the non-adhesins were qualified with a digit '0'.

30 During predictions, the network is fed with new data from the sequences that were not part of training set. Each network assigns a probability value of being an adhesin to a given sequence. The final probability is computed as described in the next section.

Probability of being an adhesin, the P_{ad} value

Query proteins are processed modularly through network trained for each attribute. Thus, five probability outputs are obtained. Final prediction was computed using the following expression which is a weighted linear sum of the probabilities from five modules:

5

$$P_{ad} = \frac{(P_A * fc_A + P_C * fc_C + P_D * fc_D + P_H * fc_H + P_M * fc_M)}{(fc_A + fc_C + fc_D + fc_H + fc_M)}$$

P_i = Probability from i module,

fc_i = fraction of correlation of i module of the trained Neural Network,

Where i = A (Amino acid frequencies), C (Charge composition), D (Dipeptide frequencies), H (Hydrophobic composition), or M (Multiplet frequencies).

10

The fraction of correlation fc_i represents the fraction of total entries that were correctly predicted ($P_{i,adhesin} > 0.5$ and $P_{i, non-adhesin} < 0.5$) by the trained network on the test set used in preliminary evaluation (Charles Anderson).

Neural Network

15

A feed forward error back propagation Neural Network was used. The program was downloaded from the web site with permission from the author, Charles W. Anderson, Department of Computer Science, Colorado State University, Fort Collins, CO 80523, anderson@cs.colostate.edu

Statistical Analysis

20

All statistical procedures were carried out using Microsoft Excel (Microsoft Corporation Inc. USA).

Sequence analysis

25

Homology analysis was carried out using CLUSTAL W (Thompson *et al* 1994), BLAST (Altschul *et al* 1990), CDD (conserved domain database) search (Marchler-Bauer *et al* 2002).

The whole genome sequences of microbial pathogens present new opportunities for the development of clinical applications such as diagnostics and vaccines. The present invention provides new leads for the development of candidate genes, and their encoded proteins in their functional relevance to preventive therapeutics.

30

The protein sequences of both the classes, i.e. adhesin and non-adhesin, were downloaded from the existing database (National Centre for Biotechnology Information (NCBI), USA). A total of 105 compositional properties under the five sequence

attributes namely, amino acid composition, multiplet composition, dipeptide composition, charge composition and hydrophobic composition were computed by computer programs written in C language. The attributes were computed for all the proteins in both the databases. The sequence-based attributes were then used to train Artificial Neural Network for each of the protein attributes. Adhesins were qualified by the digit '1' and non-adhesins were qualified by the digit '0'. Finally each trained Artificial Neural Network was used to identify potential adhesins which can be envisaged to be useful for the development of preventive therapeutics against pathogenic infections. Accordingly, the invention provides a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential, which comprises:

1. preparing two comprehensive data-sets of adhesin and non-adhesin proteins from publicly available information on protein sequences,
2. calculating computationally the sequence based attributes of the protein sequences in the publicly available protein datasets using specially developed Software for Prediction of Adhesins and Adhesin-like proteins using Neural Networks (SPAAN),
3. training the Artificial Neural Network (ANN) for the selected attributes,
4. assigning probability value suitable for an adhesin, " P_{ad} " to the query protein and identifying adhesin like property in the query proteins with the help of trained Artificial Neural Network implemented in SPAAN,
5. validating computationally the protein sequences as therapeutic potentials by comparing with the known protein sequences that are biochemically characterized in the pathogen genome.

In an embodiment of the invention the protein sequence data may be taken from an organism, specifically but not limited to organisms such as *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum*, Severe Acute Respiratory Syndrome associated coronavirus.

In another embodiment to the present invention different sequence-based attributes used for identification of proteins of therapeutic potential, comprise amino acid composition, charge composition, hydrophobicity composition, multiplets frequencies, and dipeptide frequencies.

In an embodiment, the non-homologous adhesin protein sequence may be compared with that of known sequences of therapeutic applications in the selected pathogens.

In an embodiment of the invention, the sequences of adhesin or adhesin like proteins comprise sequences of sequences IDs listed in Tables 5 and 6 identified by the method
5 of invention.

.Another embodiment of the invention the computer system comprises a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on which the
10 central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

In One embodiment of the present invention, the particulars of the organisms such as their name, strain, accession number in NCBI database and other details are given in Table 2:

15 The invention is further explained with the help of the following examples, which are given by illustration and should be construed to limit the scope of the present invention in any manner.

Example 1

Operating SPAAN:

20 The purpose of the program is to computationally calculate various sequence-based attributes of the protein sequences.

The program works as follows:

The internet downloaded FASTA format files obtained from <http://www.ncbi.nlm.nih.gov> were saved by the name <organism_name>.faa are
25 converted in the standard format by C program and passed as input to another set of C programs which computes the 5 different attributes of protein sequences (a total of 105 compositional properties in all 5 modules).

The computed properties were fed as input to the 5 different Neural Networks. Each trained network assigns a probability value of being an adhesin for a query protein. The
30 final probability (P_{ad}) was calculated as weighted average of these five individual probabilities. The weights were determined from a correlation value of correct prediction during test runs of each of the five modules.

Input/Output format:

Downloaded Files and their format:

<organism_name>.faa: file which stores the annotation and the protein sequence.

Input file Format: FASTA

5 ">gi.vertline."<annotation>

For example,

>gi.vertline.2314605.vertline.gb.vertline.AAD08472.vertline.histidine and glutamine-rich protein

10 MAHHEQQQQQQANSQHSHHHHHHAHHHHYYGGGEHHHHNAQQHAEQQAEQQ
AQQQQQQQQAHHQQQQQKAQQQNQQY

>gi.vertline.3261822.vertline.gn1.vertline.PID.vertline.e328405 PE_PGRS

MIGDGANGGPGQPGGPGGLLYGNNGGHGGAGAAGQDRGAGNSAGLIGNGGAG
GAGGNGGIGGAGAPGGLGGDGGKGGFADEFTGGFAQGGRGGFGGNGNTGAS
GGMGGAGGAGGAGGAGGGLLIGDGGAGGAGGIGGAGGVGGGGGAGGTGGGG
15 VASAFGGGNAFGGRGGDGGDGGDGGTGGAGGARGAGGAGGAGGWLSGHSG
AHGAMGSGGEGGAGGGGGGARGEAGAGGGTSTGTNPGKAGAPGTQGDSGDP
GPPG

>gi.vertline... .

Table 1: Output file format given by SPAAN

20 <organism_name>.out

| SN | P _A | P _C | P _D | P _H | P _M | P _{ad} -value | Protein Name |
|----|----------------|----------------|----------------|----------------|----------------|------------------------|--|
| 1 | 0.05683 | 0.290803 | 0.441338 | 0.50304 | 0.029503 | 0.260485 | >gi.vertline.32454344.vert line.gb.vertline.AAP82966 .1. vertline.orfla polyprotein [SARS coronavirus Hong Kong ZY-2003] |
| 2 | 0.639235 | 0.166721 | 0.054583 | 0.935385 | 0.453498 | 0.462452 | >gi.vertline.32454345.vert line.gb.vertline.AAP82967 .1. vertline.orflab polyprotein [SARS coronavirus Hong Kong ZY-2003] |
| 3 | 0.65111 1 | 0.91150 4 | 0.43869 6 | 0.54394 4 | 0.92404 4 | 0.690247 | >gi.vertline.32454346.vert line.gb.vertline.AAP82968 .1. vertline.spike glycoprotein [SARS coronavirus Hong Kong ZY-2003] |

| | | | | | | | |
|---|----------|----------|----------|----------|----------|----------|---|
| 4 | 0.464324 | 0.655003 | 0.179503 | 0.008700 | 0.241573 | 0.300970 | >gi.vertline.32454347.vert line.gb.vertline.AAP82969 .1. vertline.Orf3a [SARS coronavirus Hong Kong ZY-2003] |
|---|----------|----------|----------|----------|----------|----------|---|

Where P_A , P_C , P_D , P_H , P_M are the outputs of the five Neural Networks.

Example 2 organisms and sequence numbers

5 Table 2: Organism Name, Accession number, Number of base pairs, Date of release and Total number of proteins analyzed

| Organism Name | Accession Number | Number of base pairs | Date of release | Total no. of proteins |
|---|------------------|----------------------|-----------------|-----------------------|
| <i>E. coli</i> O157 H7 | NC_002695 | 5498450 | 7-Mar-2001 | 5361 |
| <i>H. influenzae</i> Rd | NC_000907 | 1830138 | 30-Sep-1996 | 1709 |
| <i>H. pylori</i> J99 | NC_000921 | 1643831 | 10-Sep-2001 | 1491 |
| <i>M. pneumoniae</i> | NC_000912 | 816394 | 2-Apr-2001 | 689 |
| <i>M. tuberculosis</i> H37Rv | NC_000962 | 4411529 | 7-Sep-2001 | 3927 |
| <i>R. prowazekii</i> strain Madrid E | NC_000963 | 1111523 | 10-Sep-2001 | 835 |
| <i>P. gingivalis</i> W83 | NC_002950 | 2343476 | 9-Sep-2003 | 1909 |
| <i>S. flexneri</i> 2a str. 2457T | NC_004741 | 4599354 | 23-Apr-2003 | 4072 |
| <i>S. mutans</i> UA159 | NC_004350 | 2030921 | 25-Oct-2002 | 1960 |
| <i>S. pneumoniae</i> R6 | NC_003098 | 2038615 | 6-Sep-2001 | 2043 |
| <i>N. meningitidis</i> serogroup A strain Z2491 | NC_003116 | 2184406 | 27-Sep-2001 | 2065 |
| <i>S. pyogenes</i> MGAS8232 | NC_003485 | 1895017 | Jan 31, 2002 | 1845 |
| <i>T. pallidum</i> subsp. pallidum str. Nichols | NC_000919 | 1138011 | 7-Sep-2001 | 1036 |
| Severe Acute Respiratory Syndrome (SARS) associated coronavirus Frankfurt 1 | AY291315 | 29727 | 11-JUN-2003 | 14 |
| SARS coronavirus HSR 1 | AY323977 | 29751 | 15-OCT-2003 | 14 |

| | | | | |
|--------------------------------------|---------------|-------|-------------|----|
| <i>SARS coronavirus</i> ZJ01 | AY29702 8 | 29715 | 19-MAY-2003 | 3 |
| <i>SARS coronavirus</i> TW1 | AY29145 1 | 29729 | 14-MAY-2003 | 11 |
| <i>SARS coronavirus</i> CUHK-Su10 | AY28275 2 | 29736 | 07-MAY-2003 | 4 |
| <i>SARS coronavirus</i> Urbani | AY27874 1 | 29727 | 12-AUG-2003 | 12 |
| <i>SARS coronavirus</i> | NC_0047 18 | 29751 | 9-Sep-2003 | 29 |
| <i>SARS coronavirus</i> Tor2 | AY27411 9 | 29751 | 16-MAY-2003 | 15 |
| <i>SARS coronavirus</i> GD01 | AY27848 9 | 29757 | 18-AUG-2003 | 12 |
| <i>SARS coronavirus</i> CUHK-W1 | AY27855 4 | 29736 | 31-JUL-2003 | 11 |
| <i>SARS coronavirus</i> BJ01 | AY27848 8 | 29725 | 01-MAY-2003 | 11 |

Example 3

The multi-layered feed forward Neural Network architecture implemented in SPAAN (figure 1). A given protein sequence in FASTA format is first processed through the 5 modules A, C, D, H, and M to quantify the five types of compositional attributes. A: Amino acid composition, C: Charge composition, D: dipeptide composition of the 20 dipeptides (NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI, HR), H: Hydrophobic composition, M: Amino acid frequencies as Multiplets. The sequence shown is part of the FimH precursor (gi 5524634) of *E. coli*. Subsequently, these numerical data are input to the input neuron layer. The directions of arrows show data flow. The number of neurons chosen in the input layer was equal to the number of the numerical input vectors of each module. The network was optimally trained through minimization of error of detection based on validate set through back propagation. The details are described in the methods. Each network module assigns a probability value of the protein being an adhesin based on the corresponding attribute. The final probability of a protein sequence being an adhesin is the P_{ad} value a weighted average of the individual probabilities and the associated fraction of correlation which is a measure of the strength of the prediction.

Example 4

Performance of SPAAN assessed using a test set of 37 adhesins and 37 non-adhesins that were not part of the training set. Matthew's correlation coefficient (Mcc, plotted on

Y-axis) for all the proteins with P_{ad} values above a given threshold (plotted on X-axis) (figure 2). The Matthew's correlation is defined as:

$$Mcc = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TN + FN)(TN + FP)(TP + FN)(TP + FP)}}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

Here TPs are adhesins, TNs are non-adhesins. In general, adhesins have high P_{ad} value, whereas non-adhesins have low P_{ad} value. Thus known adhesins with P_{ad} value above a given threshold are true positives whereas known non-adhesins with P_{ad} value below the given threshold are true negatives. The sensitivity, Sn is given by $\left(\frac{TP}{TP + FN}\right)$ and

specificity, Sp is given by $\left(\frac{TN}{TN + FP}\right)$. False negatives are those cases, wherein a known adhesin had P_{ad} value lower than the chosen threshold. Similarly, a known non-adhesin with a P_{ad} value higher than the chosen threshold was taken as false positive. A theoretical polynomial curve of second order (dashed line) was fitted to the observed curve (smooth line) with a Karl-Pearson correlation coefficient $R^2 = 0.9799$. The maximum point of the theoretical curve (where first derivative vanishes and second derivative is negative) was chosen as reference (vertical dotted line) to identify the maximum $Mcc = 0.94$ on the observed curve (shown by arrow). The corresponding P_{ad} value threshold was 0.51. At this P_{ad} value threshold, Sn and Sp were 0.89 and 1.0 respectively. Note that the Mcc does not drop down to the x-axis because the highest P_{ad} value attained by adhesins was 0.939 in comparison to the theoretical attainable limit of 1.0.

Example 5

Assessment of SPAAN on well known adhesins from various bacterial pathogens.

Table 3. Prediction of well characterized adhesins from various bacterial pathogens using SPAAN.

| Species | Disease caused | Adhesin ^a | Host ligand | P_{ad} value ^b (Range) |
|----------------|----------------|----------------------|--|-------------------------------------|
| <i>E. coli</i> | Diarrhoea | PapG (27) | α -D-gal(1-4) β -D-Gal-containing receptors | 0.84-0.76 |
| | | SfaS (5) | alpha-sialyl-beta-2,3-b-galactose | 0.94-0.94 |
| | | FimH (63) | D-mannosides | 0.96-0.23 ^c |

| | | | | |
|----------------------------------|--|---------------------------------------|---|------------------|
| | | Intimin (12) | tyrosine-phosphorylated form of host cell receptor Hp90 | 0.95-0.78 |
| | | PrsG (5) | Gal(alpha1-4)Gal | 0.86-0.85 |
| Nontypeable <i>H. influenzae</i> | Influenza | HMW1, HMW2 | Human epithelial cells | 0.97 |
| | | Hia (8) | human conjunctival cells | 0.93-0.90 |
| <i>H. influenzae</i> | bacterial meningitis ^d | HifE (18) | Sialylganglioside-GM1 | 0.85-0.73 |
| <i>K. pneumoniae</i> | Pneumonia | MrkD | type V collagen | 0.82 |
| <i>B. pertussis</i> | Whooping cough | FHA | Sulphated sugars on cell-surface glycoconjugates | 0.85 |
| | | Pertactin | Integrins | 0.43 |
| <i>Y. enterocolitica</i> | Enterocolitis | YadA (5) | β_1 integrins | 0.88-0.79 |
| <i>S. mutans</i> | Dental Caries | SpaP (2) | Salivary glycoprotein | 0.88, 0.87 |
| | | PAC | Salivary glycoprotein | 0.88 |
| <i>Streptococcus gordonii</i> | Oral cavity | SspA (2) | Salivary glycoprotein | 0.85, 0.84 |
| | | CshA | Fibronectin | 0.78 |
| | | CshB | Fibronectin | 0.63 |
| | | ScaA | Co-aggregation | 0.71 |
| | | SspB (2) | Salivary glycoprotein | 0.85, 0.84 |
| <i>Streptococcus sobrinus</i> | Tooth decay | SpaA | Salivary glycoprotein | 0.89 |
| | | PAG (2) | Salivary glycoprotein | 0.89, 0.73 |
| <i>Streptococcus pyogenes</i> | Scarlet Fever | Protein F | Fibronectin | 0.49 |
| <i>Streptococcus pneumoniae</i> | Bacterial Pneumonia | PsaA (5) | Human nasopharyngeal cells | 0.82-0.78 |
| | | CbpA ^e / SpsA / PbcA/ PspC | phosphorylcholine of the teichoic acid. | 0.81-0.49 |
| | | | | |
| <i>Streptococcus parasanguis</i> | Valve endocarditis | FimA | Salivary glycoprotein fibrin | 0.76 |
| <i>Streptococcus sanguis</i> | Tooth Decay | SsaB | Salivary glycoprotein | 0.71 |
| <i>Enterococcus faecalis</i> | Empyema in patients with liver disease | EfaA | Unknown | 0.83 |
| <i>Staphylococcus aureus</i> | Food Poisoning | FnbA | Fibronectin | 0.8 |
| | | FnbB (3) | Fibronectin | 0.78, 0.77, 0.69 |
| <i>Helicobacter pylori</i> | Peptic Ulcers | BabA (17) | difucosylated Lewis ^b blood group antigen | 0.87-0.68 |

^a: The number of sequences from different strains and homologs from related species analyzed are shown in parantheses.

^b: Rounded off to the second decimal.

^c: Out of 63 FimH proteins, 54 were from *E. coli*, 6 from *Shigella flexneri*, 2 from *Salmonella enterica* and 1 was from *Salmonella typhimurium*. Except 2 FimH proteins, the rest had $P_{ad} = 0.51$. The 2 exceptions (gi numbers: 5524636, 1778448) were from *E. coli*. The gi:5524636 protein is annotated as a FimH precursor but is much shorter (129 amino acids) than other members of the family. The gi:1778448 protein is a *S. typhimurium* homolog in *E. coli*.

^d: Other ailments include pneumonia, epiglottitis, osteomyelitis, septic arthritis and sepsis in infants and older children.

^e: The adhesin CbpA is also known by alternative names SpsA, PbcA and PspC. A total of seven sequences were analyzed. Except 1 PspC sequence, the rest all had $P_{ad} = 0.51$.

Example 6

Ability of SPAAN to discriminate adhesins from non-adhesins at $P_{ad} = 0.51$ (figure 3-a).

Example 7

The non-homology character of SPAAN assesses in both adhesins and non-adhesins (figure 3b and 3c).

Figure 3 (a – c). SPAAN is non-homology based software. A total of 130 adhesins and 130 non-adhesins were analyzed to assess whether the predictive power of SPAAN could be influenced by the sequence relationships. (a) Histogram plots of the number of proteins in the various P_{ad} value ranges are shown. Shaded bars represent adhesins whereas open bars represent non-adhesins. Note the SPAAN's ability to segregate adhesins and non-adhesins into two distinct cohesive groups. (b) Pairwise sequence relationships among the adhesins were determined using CLUSTAL W and plotted on X-axis. Higher scores indicate similar pairs. The corresponding differences in P_{ad} values in the same protein pair was plotted on the Y-axis. Each point in the diagram represents a pair. Arrow points to protein pairs of the FimH family with high ΔP_{ad} values in spite of high similarity: Since one of the FimH proteins (gi: 5524636) had very low P_{ad} value all pairs with this false negative protein show high ΔP_{ad} values. The protein (gi: 5524636) is of much shorter length compared with other members of the same family. (c) plot for non-adhesins. Data are plotted in the 4 quadrant format for clear inspection. Note that among protein pairs with CLUSTAL W score < 20 the majority (82% in adhesins and 86% in non-adhesins) have $\Delta P_{ad} < 0.2$. These data support the non-homology character of SPAAN.

Example 8

Genomescan of pathogens by SPAAN identifies well known adhesins and new adhesins and adhesin-like proteins

Table 4. Analysis of predictions made by SPAAN on genome scans of a few selected
5 pathogenic organisms^a

| Protein Class | Species | <i>Escherichia coli</i> O157:H7 | <i>Mycobacterium tuberculosis</i> H37Rv | SARS associated corona virus (11 strains) |
|--|---------|---------------------------------|---|---|
| Total number of proteins with P _{ad} 0.51 | | 575 | 435 | 5 |
| Known adhesins | | 17 ^b | - | - |
| Putative proteins with adhesin like characteristics | | 92 ^c | 105 ^j | - |
| Hypothetical proteins with adhesin-like characteristics | | 22 ^d | - | - |
| Proteins likely to be extracytoplasmic or located at surface | | 190 ^e | 191 ^k | 5 ^m |
| Phage proteins | | 30 ^f | - | - |
| Others | | 13 ^g | 6 ^l | - |
| Hypothetical proteins | | 157 ^h | 86 ^h | - |
| Wrong predictions | | 54 ⁱ | 47 ^l | - |

^a: SPAAN has general applicability. The three pathogens chosen here are those in which intense investigations are being conducted presently. *M. tuberculosis* is of special importance to developing countries.

10 ^b: Fimbrial adhesins, AidA-I, gamma intimin, curlin, translocated intimin receptor, putative adhesin and transport, Iha, prepilin peptidase dependent protein C.

^c: These proteins have been annotated as proteins with a putative function. These sequences were analyzed using CDD (Conserved domain database, NCBI) and BLAST searches. Adhesin like domains were found in these proteins.

15 ^d: These proteins have been annotated as 'hypothetical'. These sequences were analyzed using CDD and BLAST searches. Adhesin like domains were found in these proteins.

^e: These proteins are outer membrane, extracellular, transport, surface, exported, flagellar, periplasmic lipoprotein, and proteins annotated as 'hypothetical' but found to have similar functions listed here using BLAST and CDD searches.

5 ^f: The phage proteins were of the following functional roles – tail fiber, head decoration, DNA injection, tail, major capsid, host specificity, endolysin.

^g: Proteins predicted by SPAAN but not readily classifiable into the classes listed here have been collectively grouped as 'Others'. However, some of these proteins are known to participate in host-pathogen interactions. The annotated functional roles are typeIII secretion, antibiotic resistance, heat shock, acid shock, structural, tellurium resistance, terminase, Hcp-like, Sec-independent translocase, uncharacterized nucleoprotein, HicB-like.

^h: These proteins have been annotated as hypothetical. Re-analyses of these proteins using BLAST and CDD failed identify any function for these proteins.

15 ⁱ: These proteins have been annotated with functional roles that are very likely to occur within the cell. Hence these proteins may have remote possibility of functioning as adhesins or adhesin-like proteins. Therefore this set of proteins have been incorrectly predicted as adhesins or adhesin-like by SPAAN.

^j: These proteins are PE_PGRS, PE proteins. Several reports (for example Brennan *et al.*) indicate that PE_PGRS proteins may be localized to cell surface and aid in host-pathogen interaction.

20 ^k: Lipoproteins (lpp, lpq, lpr), PPE, outer membrane, surface, transport, secreted, periplasmic, extracellular, ESAT-6, peptidoglycan binding, exported, mpt (with extracellular domains), and proteins annotated as 'hypothetical' but found to have similar functions listed here using BLAST and CDD searches.

25 ^l: These proteins were of the following functions - glutaredoxin-like thioltransferase, putative involvement in molybdate uptake, ATP synthase chain, sulphotransferases, S.erythraea rhodanese-like protein M29612|SERCYSA_5, unknown function.

^m: These proteins were the spike glycoprotein with antigenic properties, and nsp2, nsp5, nsp6 and nsp7.

Table 5: New putative adhesins predicted by SPAAN in the genomes listed in table**2 –****(Total number = 279)**

| Protein GI Number | Gene ID | Protein name |
|---------------------------------|---------|--|
| <i>Escherichia coli</i> O157:H7 | | |
| 13360742 | 912619 | hemagglutinin/hemolysin-related protein |
| 13362986 | 914770 | putative ATP-binding component of a transport system |
| 13361114 | 913228 | putative tail fiber protein |
| 13364757 | 913676 | minor fimbrial subunit/D-mannose specific adhesin |
| 13362687 | 915687 | putative fimbrial-like protein |
| 13360856 | 912599 | AidA-I adhesin-like protein |
| 13364140 | 915374 | putative fimbrial protein |
| 13359793 | 914435 | putative invasin |
| 13364768 | 913650 | putative invasin |
| 13364034 | 915471 | Gamma intimin |
| 13362703 | 915668 | putative DNA transfer protein precursor |
| 13364141 | 915376 | putative fimbrial protein |
| 13359819 | 914463 | AidA-I adhesin-like protein |
| 13360480 | 917768 | putative fimbrial-like protein |
| 13362692 | 915681 | putative fimbrial-like protein |
| 13362585 | 916824 | putative ATP-binding component of a transport system |
| 13359881 | 914526 | putative flagellin structural protein |
| 13361579 | 917311 | putative type 1 fimbrial protein precursor |
| 13360880 | 913991 | curlin major subunit CsgA |
| 13364036 | 915465 | translocated intimin receptor Tir |
| 13360740 | 912615 | putative major pilin protein |
| 13361582 | 917317 | putative ATP-binding component of a transport system and adhesin protein |
| 13364754 | 913683 | export and assembly outer membrane protein of type 1 fimbriae |
| 13360484 | 917767 | homolog of Salmonella FimH protein |

| | | |
|----------|--------|---|
| 13364751 | 913688 | major type 1 subunit fimbrin |
| 13359597 | 913742 | putative fimbrial protein |
| 13362550 | 916787 | putative ATP-binding component of a transport system |
| 13359595 | 913739 | putative fimbrial protein |
| 13359599 | 913748 | probable outer membrane porin protein involved in fimbrial assembly |
| 13363900 | 915704 | putative fimbrial protein precursor |
| 13361575 | 917307 | putative fimbrial-like protein |
| 13364756 | 913678 | fimbrial morphology |
| 13363496 | 916142 | truncated putative fimbrial protein |
| 13359601 | 913761 | putative fimbrial-like protein |
| 13364145 | 915368 | putative type 1 fimbrial protein |
| 13363902 | 915708 | putative outer membrane usher protein precursor |
| 13361576 | 917309 | putative outer membrane protein |
| 13361013 | 913353 | putative major tail subunit |
| 13364755 | 913682 | fimbrial morphology |
| 13360738 | 912793 | putative outer membrane usher protein |
| 13363928 | 915608 | alpha-amylase |
| 13363495 | 916144 | putative outer membrane protein |
| 13362383 | 916617 | putative type-1 fimbrial protein |
| 13364373 | 914972 | outer membrane vitamin B12 receptor protein BtuB |
| 13360879 | 912479 | minor curlin subunit precursor CsgB |
| 13360739 | 912756 | putative chaperone protein |
| 13361574 | 917314 | putative fimbrial-like protein |
| 13361127 | 913212 | outer membrane protease precursor |
| 13363210 | 916442 | putative lipoprotein |
| 13361104 | 913238 | major tail protein |
| 13361709 | 917446 | putative major tail subunit |
| 13359725 | 914366 | outer membrane pore protein PhoE |
| 13360875 | 913765 | curli production assembly/transport component CsgF |
| 13362170 | 913927 | putative outer membrane protein |
| 13361473 | 917203 | putative BigB-like protein |

| | | |
|----------------------------------|--------|---|
| 13364025 | 915286 | EspF protein |
| 13360081 | 916982 | outer membrane receptor for ferric enterobactin (enterochelin) and colicins B and D |
| 13362977 | 914779 | hypothetical lipoprotein |
| 13360351 | 917632 | outer membrane protein X |
| 13360696 | 914208 | putative outer membrane precursor |
| 13361456 | 917206 | putative outer membrane protein |
| 13361626 | 917374 | putative outer host membrane protein precursor |
| 13361698 | 917449 | putative outer membrane protein |
| 13362186 | 913421 | putative outer membrane protein precursor |
| 13362697 | 915676 | long-chain fatty acid transport protein FadL |
| 13360918 | 914188 | flagellar hook protein FlgE |
| 13360737 | 912506 | putative outer membrane protein |
| 13360342 | 917629 | putative outer membrane receptor for iron transport |
| 13363396 | 916248 | outer membrane channel TolC |
| 13361958 | 912705 | putative scaffolding protein in the formation of a murein-synthesizing holoenzyme |
| 13359921 | 914566 | nucleoside-specific channel-forming protein TSX |
| 13360944 | 913890 | outer membrane receptor for ferric iron uptake |
| 13359998 | 914644 | putative outer membrane transport protein |
| 13363390 | 916251 | putative ferrichrome iron receptor precursor |
| 13364227 | 915153 | outer membrane phospholipase A |
| 13361982 | 912846 | putative outer membrane protein |
| 13360129 | 917032 | a minor lipoprotein |
| 13361817 | 912692 | putative outer membrane protein |
| 13360233 | 917507 | membrane spanning protein TolA |
| 13362837 | 915218 | putative outer membrane lipoprotein |
| 13362328 | 912985 | putative colanic acid biosynthesis glycosyl transferase |
| <i>Haemophilus influenzae</i> Rd | | |
| 16272254 | 949521 | prepilin peptidase-dependent protein D |
| 16272928 | 950762 | immunoglobulin A1 protease |
| 16272129 | 951072 | lipoprotein |

| | | |
|--------------------------------|--------|--|
| 16273251 | 950616 | hemoglobin-binding protein |
| 30995429 | 950130 | opacity protein |
| 16272854 | 949634 | protective surface antigen D15 |
| 16272283 | 950648 | opacity associated protein |
| 16272604 | 949701 | hemoglobin-binding protein |
| <i>Helicobacter pylori</i> J99 | | |
| 4155101 | 889167 | putative vacuolating cytotoxin (VacA) paralog |
| 4154798 | 890022 | putative vacuolating cytotoxin (VacA) paralog |
| 4155426 | 890036 | putative vacuolating cytotoxin (VacA) paralog |
| 4155390 | 890075 | vacuolating cytotoxin |
| 4155400 | 890058 | outer membrane protein - adhesin |
| 4155681 | 889718 | putative Outer membrane protein |
| 4155420 | 890042 | Outer membrane protein/porin |
| 4155775 | 889799 | outer membrane protein - adhesin |
| 4155419 | 890044 | Outer membrane protein/porin |
| 4154526 | 889066 | putative Outer membrane protein |
| 4154724 | 889419 | putative Outer membrane protein |
| 4155862 | 890404 | putative Outer membrane protein |
| 4156048 | 889958 | putative IRON(III) DICITRATE TRANSPORT PROTEIN |
| 4154510 | 889297 | putative Outer membrane protein |
| 4155432 | 889515 | putative outer membrane protein |
| 4155623 | 889671 | putative Outer membrane protein |
| 4155700 | 889739 | putative Outer membrane function |
| 4154740 | 889426 | Outer membrane protein/porin |
| 4155692 | 889743 | putative Outer membrane protein |
| 4155594 | 889648 | putative outer membrane protein |
| 4155680 | 889719 | putative Outer membrane protein |
| 4155217 | 890243 | putative Outer membrane protein |
| 4155958 | 889905 | putative Outer membrane protein |
| 4155201 | 890259 | putative Outer membrane protein |
| 4155013 | 889232 | cag island protein |
| 4154974 | 889032 | putative Outer membrane protein |

| | | |
|---------|--------|--|
| 4155214 | 890244 | putative Outer membrane protein |
| 4154973 | 889042 | Outer membrane protein |
| 4155344 | 890115 | putative Outer membrane protein |
| 4155099 | 889160 | FLAGELLIN A |
| 4155023 | 888978 | cag island protein |
| 4155035 | 889201 | cag island protein, CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN |
| 4155289 | 890164 | NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR |

Mycoplasma pneumoniae

| | | |
|----------|--------|--|
| 13507881 | 877207 | involved in cytodherence |
| 13507880 | 877268 | ADP1_MYCPN adhesin P1 |
| 13508228 | 877211 | species specific lipoprotein |
| 13508181 | 877124 | species specific lipoprotein |
| 13508179 | 877071 | Mollicute specific lipoprotein, MG307 homolog, from M. genitalium |
| 13508178 | 877118 | Mollicute specific lipoprotein, MG307 homolog, from M. genitalium, |
| 13508176 | 876797 | Mollicute specific lipoprotein, MG307 homolog, from M. genitalium |
| 13508175 | 876848 | Mollicute specific lipoprotein, MG307 homolog, from M. genitalium |
| 13508106 | 876953 | involved in cytodherence |
| 13508350 | 877112 | similar to phosphate binding protein Psts |

Mycobacterium tuberculosis H37 Rv

| | | |
|----------|--------|---------|
| 15607496 | 886491 | PPE |
| 15607445 | 886592 | PPE |
| 15610644 | 888270 | PE_PGRS |
| 15608588 | 886605 | PE_PGRS |
| 15609627 | 887941 | PE_PGRS |
| 15610643 | 888256 | PE_PGRS |
| 15607718 | 887725 | PE_PGRS |

| | | |
|----------|--------|----------------|
| 15609054 | 885362 | PPE |
| 15610486 | 888113 | PPE |
| 15610483 | 888120 | PPE |
| 15610479 | 888033 | PPE |
| 15609771 | 888573 | PE_PGRS |
| 15610648 | 888306 | PE_PGRS |
| 15610481 | 888114 | PE_PGRS |
| 15608117 | 885264 | PE_PGRS |
| 15607973 | 885391 | PE_PGRS |
| 15608231 | 885258 | PE_PGRS |
| 15608906 | 885429 | PE_PGRS |
| 15608891 | 885544 | PPE |
| 15609990 | 888171 | PE_PGRS |
| 15609055 | 885506 | PPE |
| 15608227 | 887094 | PE_PGRS |
| 15610524 | 888151 | PE_PGRS |
| 15609490 | 886003 | PPE |
| 15607886 | 888664 | PE_PGRS |
| 15609624 | 887909 | PE_PGRS |
| 15607420 | 886621 | PE_PGRS |
| 15608897 | 885325 | PE_PGRS(wag22) |
| 15608590 | 886595 | PE_PGRS |
| 15609728 | 887992 | PE_PGRS |
| 15608012 | 885742 | PE_PGRS |
| 15608534 | 886745 | PE_PGRS |
| 15608940 | 885730 | PE_PGRS |
| 15607887 | 888662 | PE_PGRS |
| 15609235 | 888312 | PE_PGRS |
| 15610694 | 887822 | PPE |
| 15609533 | 885517 | PE_PGRS |
| 15610480 | | PE_PGRS |

Rickettsia prowazekii strain Madrid E

| | | |
|----------|--------|-----------------------------|
| 15604316 | 883411 | CELL SURFACE ANTIGEN (sca3) |
|----------|--------|-----------------------------|

| | | |
|----------|--------|-----------------------------|
| 15604546 | 883694 | CELL SURFACE ANTIGEN (sca5) |
|----------|--------|-----------------------------|

Porphyromonas gingivalis W83

| | | |
|----------|---------|----------------------------|
| 34541453 | 2551934 | hemagglutinin protein HagA |
|----------|---------|----------------------------|

| | | |
|----------|---------|-----------------------|
| 34540040 | 2551409 | lipoprotein, putative |
|----------|---------|-----------------------|

| | | |
|----------|---------|----------------------------------|
| 34540364 | 2552375 | extracellular protease, putative |
|----------|---------|----------------------------------|

| | | |
|----------|---------|----------------------------|
| 34541613 | 2552074 | hemagglutinin protein HagE |
|----------|---------|----------------------------|

| | | |
|----------|---------|----------------------------|
| 34540183 | 2551891 | internalin-related protein |
|----------|---------|----------------------------|

Shigella flexneri 2a str. 2457T

| | | |
|----------|---------|--|
| 30065424 | 1080663 | minor fimbrial subunit, D-mannose specific adhesin |
|----------|---------|--|

| | | |
|----------|---------|---|
| 30062726 | 1077662 | putative adhesion and penetration protein |
|----------|---------|---|

| | | |
|----------|---------|--------------------------------|
| 30063758 | 1078834 | putative fimbrial-like protein |
|----------|---------|--------------------------------|

| | | |
|----------|---------|--------------------------------------|
| 30065431 | 1080671 | major type 1 subunit fimbrin (pilin) |
|----------|---------|--------------------------------------|

| | | |
|----------|---------|------------------------|
| 30063366 | 1078379 | flagellar protein FlhD |
|----------|---------|------------------------|

| | | |
|----------|---------|---------------------------------|
| 30064308 | 1079668 | outer membrane fluffing protein |
|----------|---------|---------------------------------|

| | | |
|----------|---------|-----------------------------|
| 30062613 | 1077555 | flagellar hook protein FlgE |
|----------|---------|-----------------------------|

| | | |
|----------|---------|------------------------------------|
| 30061954 | 1076843 | conserved hypothetical lipoprotein |
|----------|---------|------------------------------------|

| | | |
|----------|---------|-----------------|
| 30065173 | 1080393 | putative lipase |
|----------|---------|-----------------|

| | | |
|----------|---------|---|
| 30065425 | 1080664 | minor fimbrial subunit, precursor polypeptide |
|----------|---------|---|

| | | |
|----------|---------|---------------------------|
| 30064485 | 1079637 | putative fimbrial protein |
|----------|---------|---------------------------|

| | | |
|----------|---------|--|
| 30062615 | 1077558 | flagellar basal body L-ring protein FlgH |
|----------|---------|--|

| | | |
|----------|---------|---------------------------------|
| 30064307 | 1079452 | outer membrane fluffing protein |
|----------|---------|---------------------------------|

| | | |
|----------|---------|--------------------------------|
| 30065601 | 1080859 | putative glycoprotein/receptor |
|----------|---------|--------------------------------|

| | | |
|----------|---------|--------------------------------|
| 30062118 | 1077025 | putative fimbrial-like protein |
|----------|---------|--------------------------------|

| | | |
|----------|---------|-------------|
| 30064099 | 1079223 | lipoprotein |
|----------|---------|-------------|

| | | |
|----------|---------|--|
| 30062616 | 1077559 | flagellar basal body P-ring protein FlgI |
|----------|---------|--|

| | | |
|----------|---------|--------------------------------|
| 30063546 | 1078596 | putative fimbrial-like protein |
|----------|---------|--------------------------------|

| | | |
|----------|---------|---------------------------------|
| 30062940 | 1077910 | putative outer membrane protein |
|----------|---------|---------------------------------|

| | | |
|----------|---------|---|
| 30065426 | 1080665 | minor fimbrial subunit, precursor polypeptide |
|----------|---------|---|

| | | |
|----------|---------|---------------------------------|
| 30062779 | 1077721 | putative outer membrane protein |
|----------|---------|---------------------------------|

| | | |
|----------|---------|----------------------|
| 30064194 | 1079329 | putative lipoprotein |
|----------|---------|----------------------|

| | | |
|----------|---------|-----------|
| 30063365 | 1078378 | flagellin |
|----------|---------|-----------|

| | | |
|-----------------------------------|---------|---|
| 30062298 | 1077222 | outer membrane protein X |
| 30064968 | 1080175 | putative major fimbrial subunit |
| 30061858 | 1076740 | outer membrane pore protein E (E,Ic,NmpAB) |
| 30062178 | 1080410 | minor lipoprotein |
| 30062479 | 1077412 | putative fimbrial-like protein |
| 30062565 | 1077506 | minor curlin subunit precursor |
| 30063880 | 1078972 | putative outer membrane lipoprotein |
| 30064531 | 1079686 | cytoplasmic membrane protein |
| 30065033 | 1080243 | putative receptor protein |
| <i>Streptococcus mutans</i> UA159 | | |
| 24378550 | 1029610 | putative secreted antigen GbpB/SagA; putative peptidoglycan hydrolase |
| 24379087 | 1028055 | cell surface antigen SpaP |
| 24380463 | 1029310 | putative membrane protein |
| 24379075 | 1028046 | penicillin-binding protein 2b |
| 24378955 | 1027967 | penicillin-binding protein 1a; membrane carboxypeptidase |
| 24379801 | 1028662 | glucan-binding protein C, GbpC |
| 24379528 | 1029536 | hypothetical protein; possible cell wall protein, WapE |
| 24379231 | 1028158 | putative glucan-binding protein D; BglB-like protein |
| 24380488 | 1029325 | conserved hypothetical protein; possible transmembrane protein |
| 24380291 | 1029139 | putative amino acid binding protein |
| 24379342 | 1028247 | putative penicillin-binding protein, class C; fmt-like protein |
| 24380047 | 1028904 | putative ABC transporter, branched chain amino acid-binding protein |
| 24378698 | 1029755 | putative ABC transporter, metal binding lipoprotein; surface adhesin precursor; saliva-binding protein; lipoprotein receptor LraI (LraI family) |
| 24378708 | 1029768 | putative transfer protein |
| 24379427 | 1028331 | cell wall-associated protein precursor WapA |
| 24379272 | 1028196 | putative amino acid transporter, amino acid-binding protein |
| 24379641 | 1028511 | putative ABC transporter, amino acid binding protein |

Streptococcus pneumoniae R6

| | | |
|----------|--------|--|
| 15902395 | 934801 | Choline-binding protein |
| 15902381 | 934810 | Choline-binding protein F |
| 15902165 | 932894 | Surface protein pspA precursor |
| 15904047 | 934859 | Choline binding protein D |
| 15904036 | 933487 | Choline binding protein A |
| 15903986 | 933069 | Choline-binding protein |
| 15903796 | 933669 | Autolysin (N-acetylmuramoyl-L-alanine amidase) |

Neisseria meningitidis Z2491

| | | |
|----------|--------|---|
| 15794121 | 907145 | putative membrane protein |
| 15794144 | 907168 | putative surface fibril protein |
| 15793284 | 906275 | truncated pilin |
| 15793460 | 906456 | IgA-specific serine endopeptidase |
| 15793282 | 906273 | fimbrial protein precursor (pilin) |
| 15793337 | 906332 | adhesin |
| 15793253 | 906243 | putative lipoprotein |
| 15794356 | 907848 | putative lipoprotein |
| 15793684 | 906699 | putative membrane protein |
| 15793290 | 906281 | truncated pilin |
| 15793283 | 906274 | truncated pilin |
| 15793475 | 906471 | haemoglobin-haptoglobin-utilization protein |
| 15793406 | 906401 | porin, major outer membrane protein P.I |
| 15794985 | 907333 | adhesin MafA2 |
| 15794344 | 907836 | putative lipoprotein |
| 15794622 | 908118 | hypothetical outer membrane protein |
| 15793599 | 906604 | pilus-associated protein |
| 15793763 | 906779 | putative periplasmic binding protein |

Streptococcus pyogenes MGAS8232

| | | |
|----------|--------|--|
| 19745214 | 995235 | putative secreted protein |
| 19746570 | 994224 | putative penicillin-binding protein 1a |
| 19745593 | 994771 | putative 42 kDa protein |
| 19745813 | 993958 | putative adhesion protein |

| | | |
|----------|--------|-----------------------------------|
| 19745225 | 994839 | putative choline binding protein |
| 19745828 | 995250 | streptolysin S associated protein |
| 19746229 | 995021 | putative minor tail protein |
| 19746909 | 994105 | putative laminin adhesion |
| 19745560 | 995061 | putative cell envelope proteinase |

Treponema pallidum subsp. pallidum str. Nichols

| | | |
|----------|---------|-------------------------------|
| 15639714 | 2611034 | flagellar hook protein (flgE) |
| 15639609 | 2611657 | tpr protein J (tprJ) |
| 15639111 | 2610909 | tpr protein C (tprC) |
| 15639125 | 2610968 | tpr protein D (tprD) |

SARS coronavirus

| | | |
|----------|---------|--|
| 31581505 | | spike protein S [SARS coronavirus Frankfurt 1] |
| 32187357 | | spike protein S [SARS coronavirus HSR 1] |
| 32187342 | | spike glycoprotein [SARS coronavirus ZJ01] |
| 30698329 | | putative spike glycoprotein S [SARS coronavirus TW1] |
| 30421454 | | putative spike glycoprotein [SARS coronavirus CUHK-Su10] |
| 30027620 | | S protein [SARS coronavirus Urbani] |
| 29836496 | 1489668 | E2 glycoprotein precursor; putative spike glycoprotein [SARS coronavirus] |
| 30795145 | | spike glycoprotein [SARS coronavirus Tor2] |
| 31416295 | | spike glycoprotein S [SARS coronavirus GD01] |
| 30023954 | | putative E2 glycoprotein precursor [SARS coronavirus CUHK-W1] |
| 30275669 | | spike glycoprotein S [SARS coronavirus BJ01] |
| 29837498 | | 3C-like proteinase nsp5-pp1a/pp1ab (3CL-PRO) [SARS coronavirus] |
| 29837501 | | putative nsp8-pp1a/pp1ab [SARS coronavirus] |
| 29837503 | | putative nsp10-pp1a/pp1ab; formerly known as growth-factor-like protein [SARS coronavirus] |
| 29837502 | | putative nsp9-pp1a/pp1ab [SARS coronavirus] |

Table 6: Hypothetical proteins predicted as putative adhesins by SPAAN in the genomes listed in table 2 –

(Total number of proteins = 105)

Protein GI Gene ID
number

Escherichia coli O157:H7

| | |
|----------|--------|
| 13363955 | 915578 |
| 13360000 | 914929 |
| 13362244 | 912369 |
| 13359999 | 914888 |
| 13361583 | 917316 |
| 13361172 | 913156 |
| 13361131 | 913207 |
| 13359780 | 914422 |
| 13360571 | 912499 |
| 13362197 | 912893 |
| 13362260 | 912399 |
| 13360947 | 913505 |
| 13361464 | 917196 |
| 13361635 | 917367 |
| 13362421 | 916655 |
| 13361463 | 917195 |

Haemophilus influenzae Rd

| | |
|----------|--------|
| 16272115 | 951058 |
| 30995442 | 950581 |

Helicobacter pylori J99

| | |
|---------|--------|
| 4155526 | 889586 |
| 4155712 | 889748 |
| 4155632 | 889684 |
| 4156035 | 889468 |
| 4155499 | |

Mycoplasma pneumoniae

| | |
|----------|--------|
| 13507870 | 877230 |
| 13508239 | 877245 |
| 13508109 | 876868 |
| 13508025 | 877084 |
| 13507838 | 876784 |
| 13507883 | 877183 |
| 13507871 | 877239 |
| 13507944 | 877056 |
| 13508241 | 876750 |
| 13507942 | 877055 |
| 13507840 | 877387 |
| 13507867 | 877242 |
| 13508201 | 877044 |
| 13507941 | 876985 |
| 13508114 | 877397 |

Mycobacterium tuberculosis H37Rv

| | |
|----------|--------|
| 15611014 | 886198 |
| 15610173 | 887320 |
| 15609513 | 885515 |
| 15608094 | 885411 |
| 15610958 | 886155 |
| 15607528 | 886436 |
| 15607678 | 887473 |
| 15609587 | 885760 |
| 15610708 | 887227 |
| 15609526 | 885246 |
| 15611033 | 886225 |
| 15609028 | 885094 |
| 15607730 | 887771 |
| 15609121 | 885813 |
| 15608255 | 885951 |

15608409 887039

15609124 885815

15607734 887797

Rickettsia prowazekii strain Madrid E

15604649 883964

15604322 883472

15604659 883996

15604417 883217

Porphyromonas gingivalis W83

34540233 2551594

Shigella flexneri 2a str. 2457T

30062687 1077638

30062956 1080449

30063681 1078754

30065435 1080675

30063891 1078983

30063211 1078195

30065233 1080463

30064387 1079531

30062638 1077590

30065236 1080466

30061839 1076721

Streptococcus mutans UA159

24378864 1029452

24380475 1029319

24380237 1029088

24379203 1028139

24380480 1029320

24379275 1029489

24379291 1028216

24379295 1028215

24379804 1028663

24379162 1029417

24378987 1029363

24379179 1028118

24379166 1028107

24378827 1029444

24380216 1029067

Streptococcus pneumoniae R6

15902140 932867

15903446 934616

15903916 934001

15903848 933609

15902832 934332

15902372 934804

15902152 932889

Neisseria meningitidis Z2491

15793668 906680

15794714 907603

Streptococcus pyogenes MGAS8232

19747011 993608

19747024 994165

19747012 994373

19746396 995057

19746651 993824

19745883 995045

19745912 994077

Treponema pallidum subsp. *pallidum* str. Nichols

15639844 2611061

15639720 2611059

Table 7: The list of 198 adhesins found in bacteria
PapG (*E. coli*)

12837502

7407201

7407207
7407205
147096
4240529
7407203
42308
7443327
78746
18265934
26111419
26250987
26109826
26249418
13506767
42301
78745
129622
147092
13506906
7407209
147080
281926
7407199
147100
78744

SfaS (E.coli)

477910
264035
42959
134449
96425

FimH (E.coli)

26251208
26111640
5524634
29422425
5524630
29422435
29422415
10946257
29422419
11120564
29422457
11120562
29422459
5524632
29422455
29422453
29422451
29422449
29422447
29422445
29422443
29422437
29422433
29422431
29422429
29422427
29422423
29422421
29422417
729494
1361011
1790775

3599571
29422441
12620398
29422439
5524628
1787779
1742472
1742463
15801636
25321294
12515169
11120566
24051859
24112911
13360484
15800801
15830279
25392018
25500156
12514120
1787173
16128908
16501811
16759519
24051219
24112354
30040724
30062478
6650093
5524636
1778448

Intimin (E.coli)

| | |
|--|----------|
| | 17384659 |
| | 4388530 |
| | 1389879 |
| | 15723931 |
| | 4323336 |
| | 4323338 |
| | 4323340 |
| | 4323342 |
| | 4323344 |
| | 4323346 |
| | 4323348 |
| | 4689314 |
| PrsG (E.coli) | |
| | 42523 |
| | 42529 |
| | 7443328 |
| | 7443329 |
| | 1172645 |
| HMW1 (Nontypeable <i>H. influenzae</i>) | |
| | 282097 |
| HMW2 (Nontypeable <i>H. influenzae</i>) | |
| | 5929966 |
| Hia (Nontypeable <i>H. influenzae</i>) | |
| | 25359682 |
| | 25359489 |
| | 25359709 |
| | 25359628 |
| | 25359414 |
| | 25359389 |
| | 21536216 |
| | 25359445 |
| HifE (<i>H. influenzae</i>) | |

| | |
|-----------------------------------|-----------------|
| | <i>13506868</i> |
| | 13506870 |
| | 13506872 |
| | 13506874 |
| | 13506876 |
| | 3688787 |
| | 3688790 |
| | 3688793 |
| | 2126301 |
| | 1170264 |
| | 1170265 |
| | 533127 |
| | 535169 |
| | 3025668 |
| | 3025670 |
| | 3025672 |
| | 3025674 |
| | 642038 |
| MrkD (<i>K. pneumoniae</i>) | |
| | <i>127307</i> |
| FHA (<i>B. pertussis</i>) | |
| | <i>17154501</i> |
| Pertactin (<i>B. pertussis</i>) | |
| | <i>33571840</i> |
| YadA (<i>Y. enterocolitica</i>) | |
| | <i>10955604</i> |
| | 4324391 |
| | 28372996 |
| | 23630568 |
| | 32470319 |
| SpaP (<i>S. mutans</i>) | |
| | <i>26007028</i> |

| | |
|--|----------|
| | 47267 |
| PAc (<i>S. mutans</i>) | |
| | 129552 |
| SspA (<i>Streptococcus gordonii</i>) | |
| | 25990270 |
| | 1100971 |
| CshA (<i>Streptococcus gordonii</i>) | |
| | 457707 |
| CshB (<i>Streptococcus gordonii</i>) | |
| | 18389220 |
| ScaA (<i>Streptococcus gordonii</i>) | |
| | 310633 |
| SspB (<i>Streptococcus gordonii</i>) | |
| | 25055226 |
| | 3220006 |
| SpaA (<i>Streptococcus sobrinus</i>) | |
| | 546643 |
| PAg (<i>Streptococcus sobrinus</i>) | |
| | 217036 |
| | 47561 |
| Protein F (<i>Streptococcus pyogenes</i>) | |
| | 19224134 |
| PsaA (<i>Streptococcus pneumoniae</i>) | |
| | 18252614 |
| | 7920456 |
| | 7920458 |
| | 7920460 |
| | 7920462 |
| CbpA ^e / SpsA / PbcA/ PspC (<i>Streptococcus pneumoniae</i>) | |
| | 14718654 |
| | 2425109 |

| | |
|----------------------------------|----------|
| | 2576331 |
| | 2576333 |
| | 3153898 |
| | 9845483 |
| | 19548141 |
| FimA (Streptococcus parasanguis) | |
| | 97883 |
| SsaB (Streptococcus sanguis) | |
| | 97882 |
| EfaA (Enterococcus faecalis) | |
| | 493017 |
| FnbA (Staphylococcus aureus) | |
| | 120457 |
| FnbB (Staphylococcus aureus) | |
| | 581562 |
| | 21205592 |
| | 13702452 |
| BabA (Helicobacter pylori) | |
| | 13309962 |
| | 13309964 |
| | 13309966 |
| | 13309968 |
| | 13309970 |
| | 13309972 |
| | 13309974 |
| | 13309976 |
| | 13309978 |
| | 13309980 |
| | 13309982 |
| | 13309984 |
| | 13309986 |
| | 13309988 |

13309990

13309992

13309994

Advantages:

1. The method helps in discovering putative adhesins, which are of great importance in drug discoveries and preventive therapeutics.
2. The method is useful in predicting the adhesive nature of even unique proteins,
5 because it is independent of the homology of the query proteins with other proteins.
3. This method is easy to use. For calculating the output, only the amino acid sequence is required as input. No other information is required to get the information about its adhesive nature.

REFERENCES MAY BE MADE TO

1. Andrea, T.A., Kalayeh, H.(1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J. Med. Chem. 34, 2824-2836.
- 5 2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.(1990) Basic local alignment search tool. J Mol Biol. 215(3), 403-410.
3. Bassinet L, Gueirard P, Maitre B, Housset B, Gounon P, Guiso N.(2000) Role of adhesins and toxins in invasion of human tracheal epithelial cells by Bordetella pertussis. Infect Immun. 68(4), 1934-1941.
- 10 4. Bock, K., *et al.*(1985). Specificity of binding of a strain of uropathogenic Escherichia coli to Gal alpha 1----4Gal-containing glycosphingolipids. J. Biol. Chem. 260, 8545-8551.
5. Brendel, V., Bucher, P., Nourbakhsh, I.R., Edwin Blaisdell, B., and Karlin, S.(1992) Methods and algorithms for statistical analysis of protein sequences.
- 15 Proc. Natl. Acad. Sci. USA 89, 2002-2006.
6. Brennan, M.J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M., Jacobs, W.R., (2001).
7. Evidence that Mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect. Immun, 69, 7326-7333.
- 20 8. De BK, Woolfitt AR, Barr JR, Daneshvar MI, Sampson JS, Ades EW, Carlone GM. (2003) Analysis of recombinant acylated pneumococcal surface adhesin A of Streptococcus pneumoniae by mass spectrometry. Arch Biochem Biophys. 15, 419(2), 147-157.
9. Eglund PG, Du LD, Kolenbrander PE (2001) Identification of independent
- 25 Streptococcus gordonii SspA and SspB functions in coaggregation with Actinomyces naeslundii. Infect Immun. 69(12), 7512-7516
10. Finlay, B.B. and Falkow, S.(1997) Common themes in microbial pathogenicity revisited. Microbiol. Mol. Biol. Rev. 61, 136-169,
11. Fraser, C.M., Eisen, J., Fleischmann, R.D., Ketchum, K.A., Peterson, S.(2000)
- 30 Comparative genomics and understanding of microbial biology. Emerg. Infect. Dis. 6,505-6512
12. Halperin, S. A., Scheifele, D., Mills, E., Guasparini, R., Humphreys, G., Barreto, L., Smith, B.(2003) Nature, evolution, and appraisal of adverse events

- and antibody response associated with the fifth consecutive dose of a five-component acellular pertussis-based combination vaccine. *Vaccine* 21, 2298–2306.
13. Hartford O, McDevitt D, Foster TJ.(1999) Matrix-binding proteins of *Staphylococcus aureus*: functional analysis of mutant and hybrid molecules. *Microbiology*. 145 (Pt 9), 2497-2505.
 14. Hobohm, U. and Sander, C.(1995) A sequence property approach to searching protein databases. *J. Mol. Biol.* 251, 390-399.
 15. Ide T, Michgehl S, Knappstein S, Heusipp G, Schmidt MA.(2003) Differential modulation by Ca²⁺ of type III secretion of diffusely adhering enteropathogenic *Escherichia coli*. *Infect Immun.* 71(4), 1725-1732.
 16. Langermann S *et al.*(2000) Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic *Escherichia coli*. *J. Infect. Dis.* 181, 774-778.
 17. Lowe A.M., Lambert, P.A., Smith, A.W.(1995) Cloning of an *Enterococcus faecalis* endocarditis antigen: homology with adhesins from some oral streptococci. *Infect Immun.* 63, 703-706.
 18. Maurer, L., Orndorff, P.(1987). Identification and characterization of genes determining receptor binding and pilus length of *Escherichia coli* type 1 pili. *J. Bacteriol.* 169, 640-645
 19. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* 1, 30(1), 281-283.
 20. Neubauer H, Hensel A, Aleksic S, Meyer H.(2000) Evaluation of a *Yersinia* adhesion gene (*yadA*) specific PCR for the identification of enteropathogenic *Yersinia enterocolitica*. *Int J Food Microbiol.* 15, 57(3), 225-227.
 21. Nishikawa, K., Kubota, Y. and Ooi, T.(1983) Classification of proteins into groups based on amino acid composition and other characters. II. grouping into four types. *J. Biochem.* 94, 997-1007.
 22. Peregrin-Alvarez, J.M., Tsoka, S., Ouzounis, C.A.(2003) The phylogenetic extent of metabolic enzymes and pathways. *Genome Res.* 13, 422-427.

23. Prinz, C., Hafsi, N. Volland, P.(2003) *Helicobacter pylori* virulence factors and the host immune response: implications for therapeutic vaccination. *Trends in Microbiol.* 11, 134-138.
24. Rapola, S., Jääntti, V., Eerola, M., Helena Mäkelä, P., Käyhty, H., Kilpi, T.(2003) Anti-PsaA and the risk of pneumococcal AOM and carriage. *Vaccine* 21, 3608–3613.
25. Rison, S.C., Teichmann, S.A., Thornton, J.M.(2002) Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in *Escherichia coli*. *J. Mol. Biol.* 318, 911-932
- 10 26. Sperandio V, Bailey C, Giron JA, DiRita VJ, Silveira WD, Vettore AL, Kaper JB.(1996) Cloning and characterization of the gene encoding the OmpU outer membrane protein of *Vibrio cholerae*. *Infect Immun.* 64(12), 5406-5409.
27. St Geme JW 3rd, Cutter D.(2000) The *Haemophilus influenzae* Hia adhesin is an autotransporter protein that remains uncleaved at the C terminus and fully cell associated. *J Bacteriol.* 182(21), 6005-6013.
- 15 28. Thompson, J.D., Higgins, D.G., Gibson, T.J.(1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680
- 20 29. Van Schilfgaarde M, van Ulsen P, Eijk P, Brand M, Stam M, Kouame J, van Alphen L, Dankert J.(2000) Characterization of adherence of nontypeable *Haemophilus influenzae* to human epithelial cells. *Infect Immun.* 68(8), 4658-4665.
30. Wizemann, T.M., Adamou, J.E., Langermann, S.(1999). Adhesins as targets for vaccine development. *Emerg. Infect. Dis.* 5, 395-403,
- 25 31. Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., and Koonin, E.V.(2001) Genome alignment, evolution of prokaryotic genome organization and prediction of gene function using genomic context. *Genome Res.* 11, 356-372
32. Yu J, Leung WK, Go MY, Chan MC, To KF, Ng EK, Chan FK, Ling TK, Chung SC, Sung JJ.(2002) Relationship between *Helicobacter pylori* babA2 status with gastric epithelial cell turnover and premalignant gastric lesions. *Gut.* 51(4), 480-484.
- 30

33. Zuegge, J., Ralph, S., Schmuker, M., McFadden, G.I., Schneider, G.(2001)
Deciphering apicoplast targeting signals--feature extraction from nuclear-
encoded precursors of Plasmodium falciparum apicoplast proteins. Gene 280,
19-26.

Claims

1. A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of:
 - a. computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, wherein the attributes are, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition,
 - b. training a artificial neural Network (ANN) for each of the computed five attributes, and
 - c. identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 .
2. A method as claimed in claim 1, wherein the protein sequences are obtained from pathogens, eukaryotes, and multicellular organisms.
3. A method as claimed in claim 1, wherein the protein sequences are obtained from the pathogens selected from a group of organisms comprising *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum* and Severe Acute Respiratory Syndrome associated human coronavirus (SARS).
4. A method as claimed in claim 1, wherein the method is a non-homology method.
5. A method as claimed in claim 1, wherein the method uses 105 compositional properties of the sequences.
6. A method as claimed in claim 1, wherein the method shows sensitivity of at least 90%.
7. A method as claimed in claim 1, wherein the method shows specificity of 100%.
8. A method as claimed in claim 1, wherein the method helps identifies adhesins from distantly related organisms.
9. A method as claimed in claim 1, wherein the neural network has multi-layer feed forward topology, consisting of an input layer, one hidden layer, and an output layer.

10. A method as claimed in claim 9, wherein the number of neurons in the input layer are equal to the number of input data points for each attribute.
11. A method as claimed in claim 1, wherein the "P_{ad}" is a weighted linear sum of the probabilities from five computed attributes.
- 5 12. A method as claimed in claim 1, wherein each trained network assigns a probability value of being an adhesin for the protein sequence.
13. A computer system for performing the method of claim 1, said system comprising a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in
10 built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.
14. A set of 274 annotated genes encoding adhesin and adhesin-like proteins,
15 having SEQ ID Nos. 385 to 658.
15. A set of 105 hypothetical genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 659 to 763.
16. A set of 279 annotated adhesin and adhesin-like proteins of SEQ ID Nos. 1 to 279.
- 20 17. A set of 105 hypothetical adhesin and adhesin-like proteins of SEQ ID Nos. 280 to 384.
18. A fully connected multilayer feed forward Artificial Neural Network based on the computational method as claimed in claim 1, comprising of an input layer, a hidden layer and an output layer which are connected in the said sequence,
25 wherein each neuron is a binary digit number and is connected to each neuron of the subsequent layer for identifying adhesin or adhesin like proteins, wherein the program steps comprise:-
 - [a] feeding a protein sequence in FASTA format;
 - [b] processing the sequence obtained in step [a] through the 5 modules
30 named A, C, D, H and M, wherein attribute A represents an amino acid composition, attribute C represents a charge composition, attribute D represents a dipeptide composition of the 20 dipeptides [NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI

and HR], attribute H represents a hydrophobic composition and attribute M represents amino acid frequencies in multiplets to quantify 5 types of compositional attributes of the said protein sequence to obtain numerical input vectors respectively for each of the said attributes wherein the sum of numerical input vectors is 105;

[c] processing of the numerical input vectors obtained in step [b] by the input neuron layer to obtain signals, wherein the number of neurons is equal to the number of numerical input vectors for each attribute;

[d] processing of signals obtained from step [c] by the hidden layer to obtain synaptic weighted signals, wherein the optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually;

[e] delivering synaptic weighted signals obtained from step [d] to the output layer for assigning of a probability value for each protein sequence fed in step [a] as being an adhesin by each network module; and

[f] using the individual probabilities obtained from step [e] for computing the final probability of a protein sequence being an adhesin denoted by the P_{ad} value, which is a weighted average of the individual probabilities obtained from step [e] and the associated fraction of correlation which is a measure of the strength of the prediction.

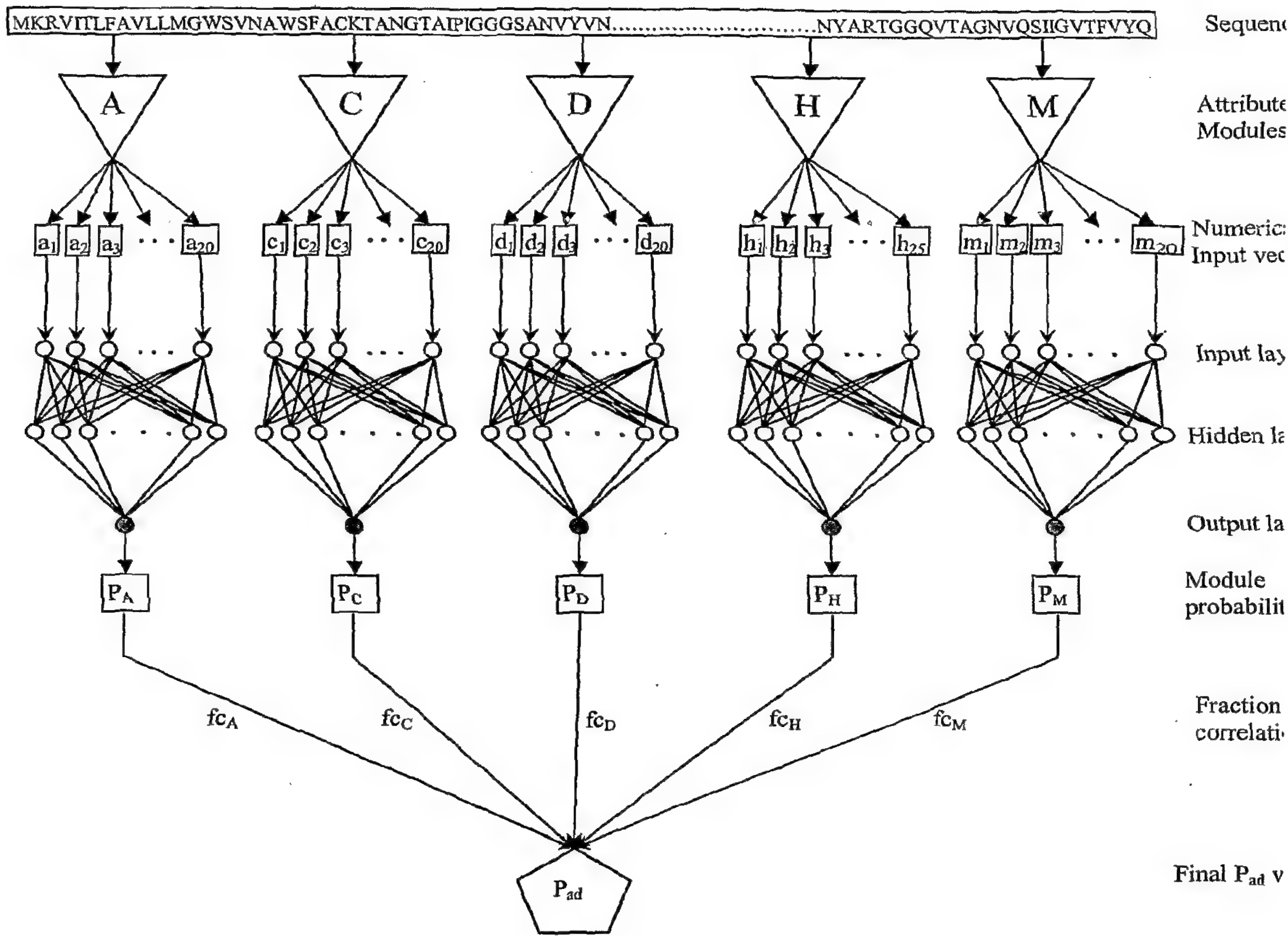
19. A network as claimed in claim 18, wherein the input neuron layer consists of a total of 105 neurons corresponding to 105 compositional properties.

20. A network as claimed in claim 18, wherein the hidden layer comprises of neurons represented as 30 for amino acid frequencies, 28 for multiplet frequencies, 28 for dipeptide frequencies, 30 for charge composition and 30 for hydrophobic composition.

21. A network as claimed in claim 18, wherein the output layer comprises of neurons to deliver the output values as probability value for each protein sequence.

The Neural Network architecture

Figure1



Assessment of SPAAN using defined test dataset.

Figure 2

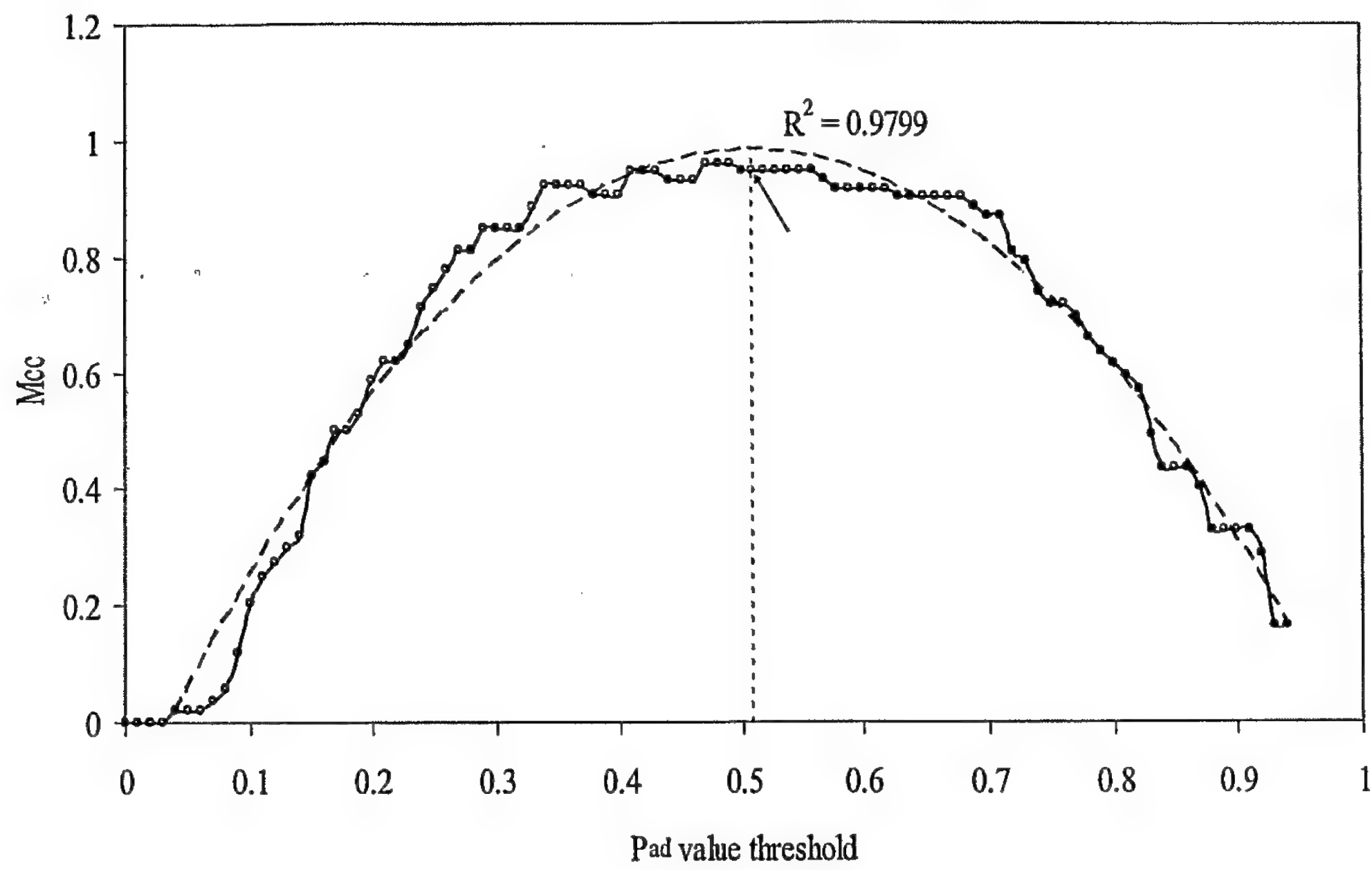


Figure 3 (a)

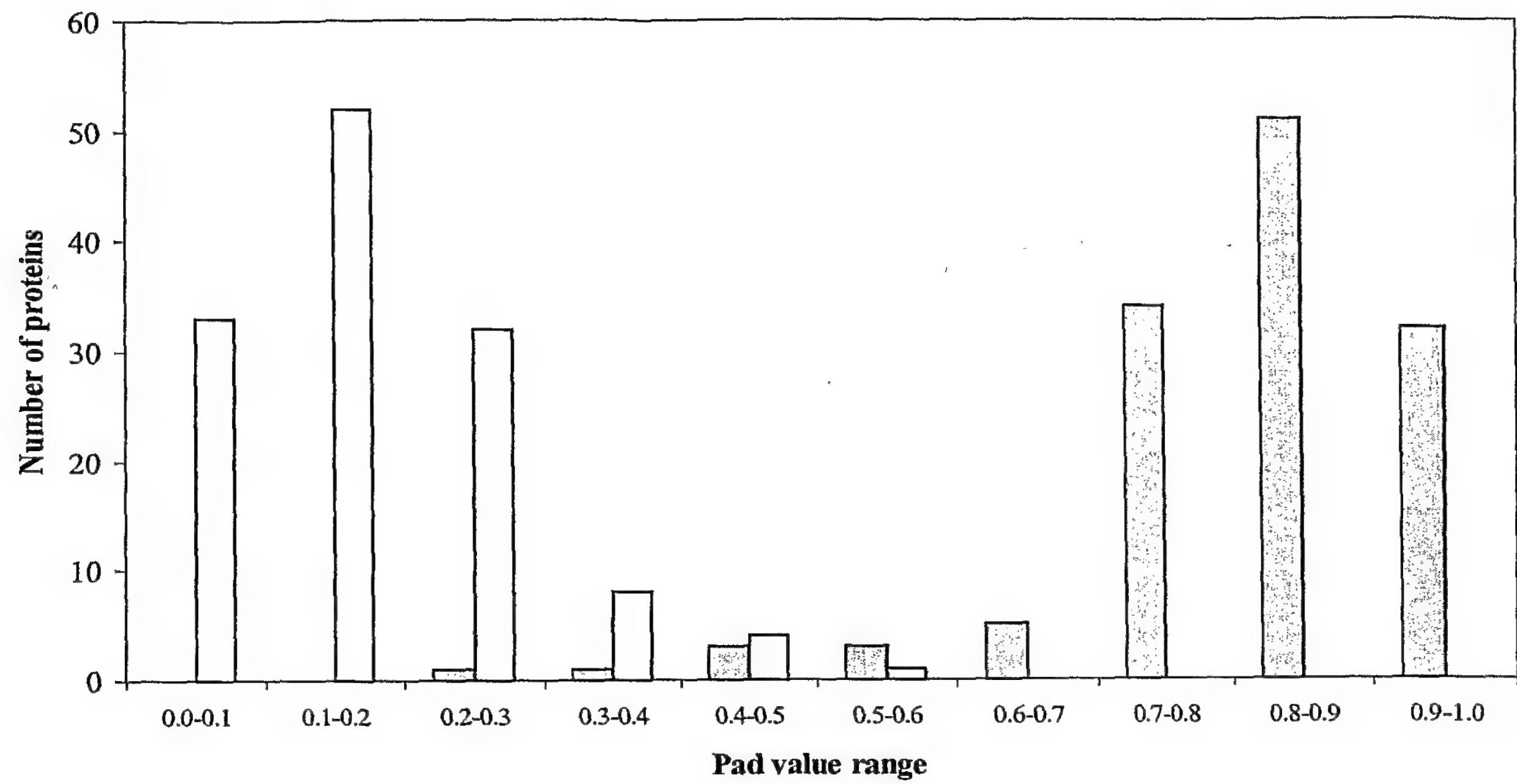


Figure 3 (b)

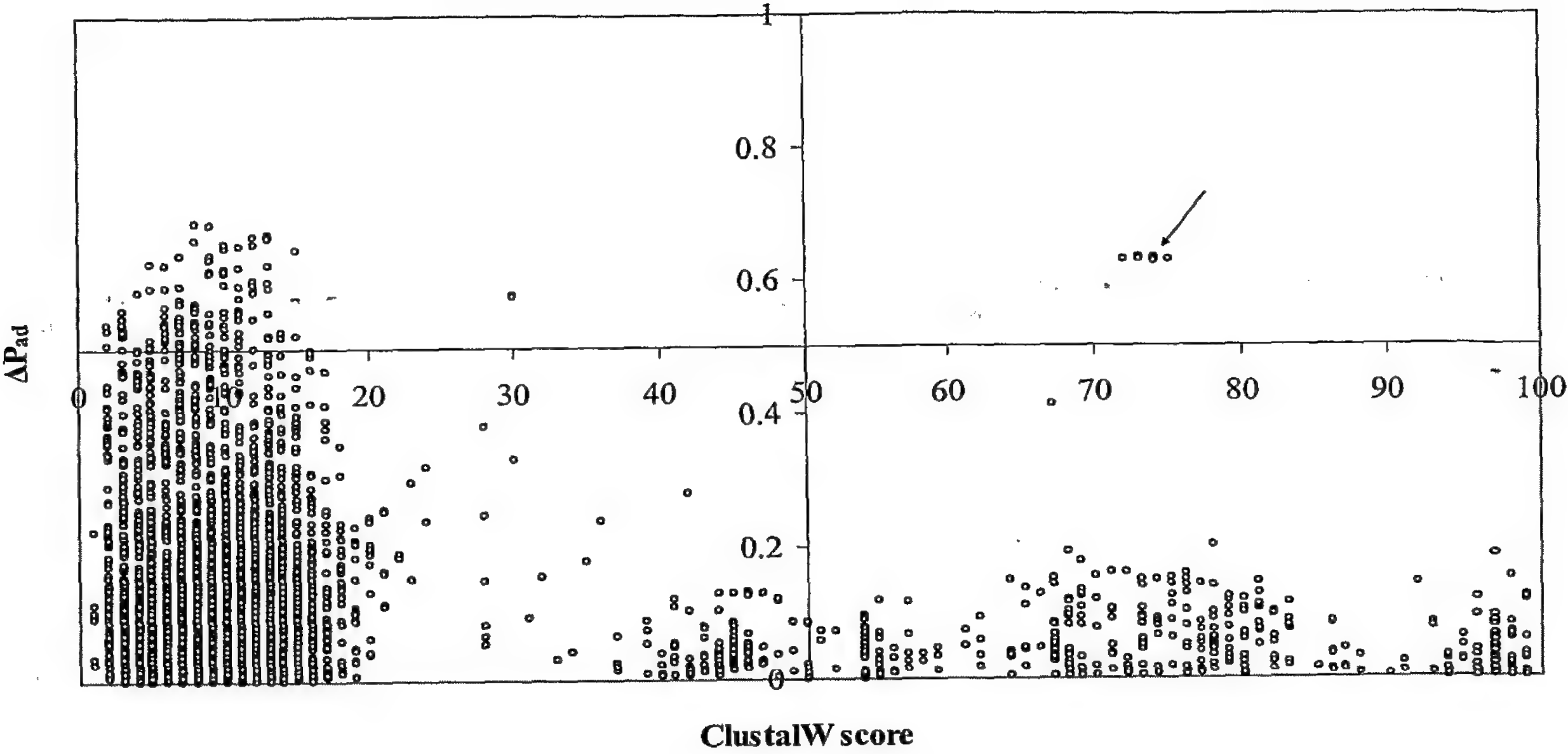
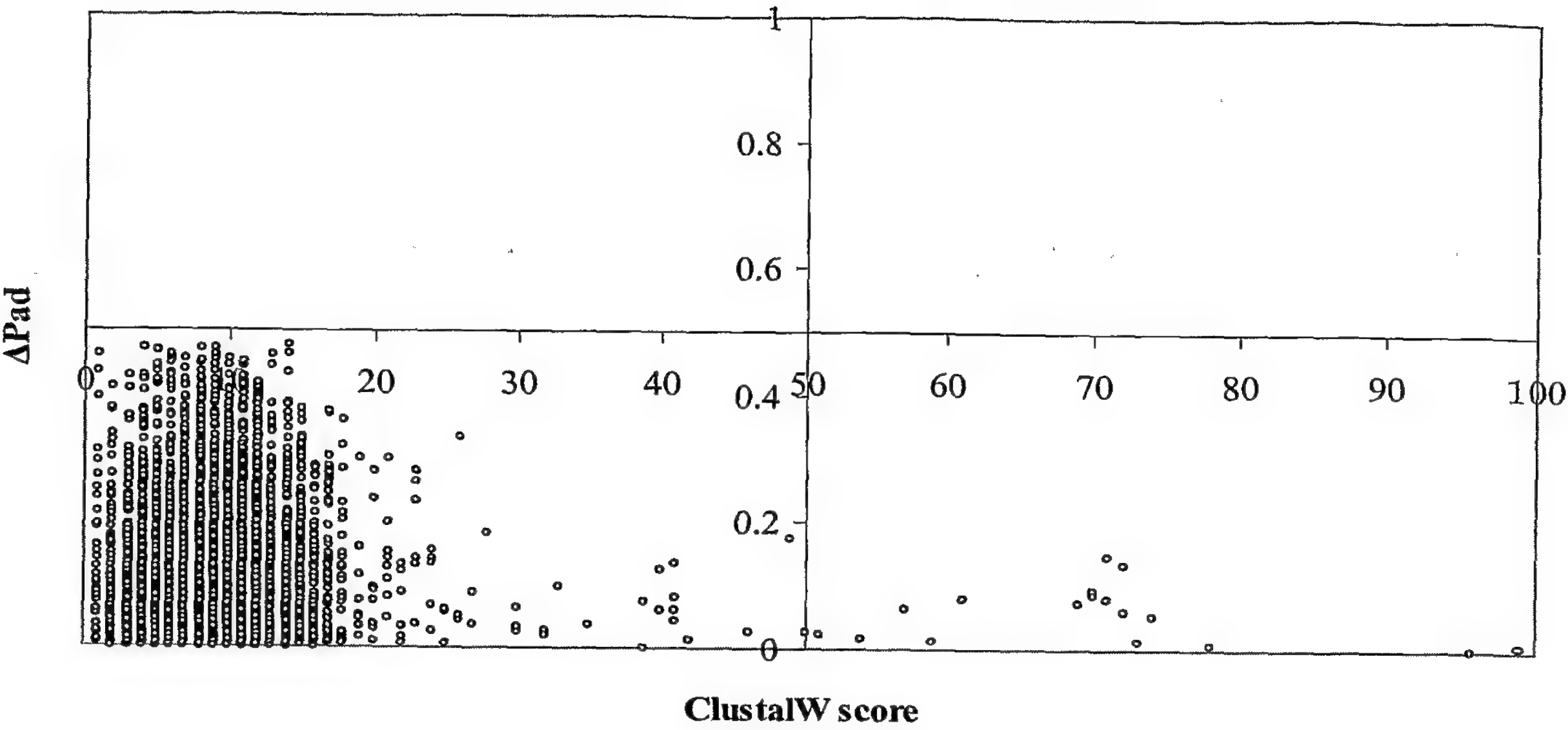


Figure 3(c)



Application Project

<120> Title :

<130> AppFileReference :

5 <140> CurrentAppNumber :

<141> CurrentFilingDate :

Sequence

10 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```
MINLSKEATV GKALTPIAIL MMLSFPVASQ AAGLVIKNGT VYNANGVPVV DINKPNGSGL      60
SHNIWDNLNV DKNGVVFNNNS ANESSTSLAG NIQGNLSNLS GSAKVILNEV TSKNPSTING    120
MMFVLDKAD 1TLANCMT VNCGGCINTC KLTLETTPPD IQDDKIDAGYS VNCSTITLGN    180
15 LDNASPTEIL SRNVVNGKV SADELNVVAG NNYVNAAGQV TGSVSATGSR NGYSVDVAKL      240
GGMYANKISL VSTEKGVGVR NLGVIAGGVN GVSIDSKGNL LNSNAQIQSA STINLTNGT      300
LDNTTGTVTS VGTISLNTNK NTIVNTRAGN ISTMGDIYVN SGTIDNTNGK LAAAGMLAVD      360
TNNATLINSK KGSSVGIEAG LVALKTGTLN NSNGQIRGGY VGLESAALNN NNGDIQTTGD      420
IAIISNGNVD NNGKLIRSST GHIVIGAAGS VNNGSTKTAD TGSSDSLGI ADTGVEIGAN      480
20 NINNNGGQIA SNGNVSLSSY STIDYAGKI LSNSKVIKKG SSLRNDTGGI SGKQGIEVAV      540
GGSLTNNIGV ISSEEGDISL LANSVDNHGG FMMGQNTIME MSGVNNNTA LIVASKKLKI      600
NARGSIENRD GNNFGNAYGL YFGMPQQTGG MVGKEGIELS GQNIYNNNSR LIAEDGPLTL      660
QAQNTFDNTR ALVTSGADAS IQVGGTYNNN YATTWSAGNL DIDATTLQNS SSGTMIDNNA      720
TGFIASDKNL SLEVNSLTN YGWISGKGDV DVTVNNGNLY NRNTIAAEKG LDIAALNGIE      780
25 NWKDISAGGD LTMNTNRHVT NNSNSNMVGQ NIVINAVNDI NNRGNIVSDA DLNVTTKGNL      840
YNYLYMVGYG DIALSANSVA NNNATIEATG DLIIDSKGNV GNNRGNLHAL NGVLSVKGN      900
LNNDNGEIRG YGDVTLALTG NYDSYKGS LT SETGDVTLTA NIVDNAYGLI AGENVSVDK      960
STIYNNTALI AANKKLIVINA GGNLENRDGN NFLRNNGALF GITDNVGGIV GKEGVTLISAQ    1020
NVYNNNSSII AENGPLNLLS RGTLDNTRAL LSSGADAIIR AAGTFYNNYA TTYSAGNLDV    1080
30 YAASLNNASD GRLEDNTATG VIASDKNLDL SVDNSVTNYG WISGKGDVHF NVLKGTLYNR    1140
NAIAADNALT INALNGVENF KDIVAGTALT IDTQKYVTNN SNSNMLGQTI AINAVNDINN    1200
RGNIVGDYSL GVKTTGNIYN YLNMLSYGVA GVSANKVTNS GKDAVLGGFY GLALEANETD    1260
NTGTIVGM                                     1268
```

<212> Type : PRT

35 <211> Length : 1268

SequenceName : SEQ ID 1

SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```
MNTIHLRCLF RMNPLVWCLW ADVAAKLRSL KRYSVFTFQR MKFMNRTSPY YCRRSVLSLL      60
ISALIYAPPG MAAFTPDVIG VVNDETVDGS QRVDERGTTN NTHIINHQQ NVYGGVSNBS    120
45 LIESGGYQDV GRHNNYVGQS NNTTINGGRQ SIHGGISTG TIIESGNQDV YKGGISNGTT    180
IKGGASRVEG GSANGTLIDG GSQIVKVQGH ADGTTINKSG SQDVVQGS LA TNTTINGGRQ    240
YVEQSTVETT TIKNGGEQRV YESRALDTTI EGGTQSLNSK STAKNTQIYS GGTQIIDNTS    300
SSDVIEVYSG GVLDVSGGTA TNVTQHDGAI LKTNTNGTTV SGTNSEGAFS IHNHVADNVL    360
LENGGHLIDN AYGSANKTII KDKGTMSVLT NAKADATRID NGGVM DVAGN ATNTIINGGT    420
50 QNINNYGIAT GTNINSGTON IKSGGKADTT IISGSRQVV EKDGTAIGSN ISAGGSLIVY    480
TGGIAHGVNQ ETGSALVANT GAGTDIEGYN KLSHFTITGG EANYVVLENT GELTVVAKTS    540
AKNTTIDAGG KLIVQKEAKT DSTRLNNGGV LEVQDGGEAK HVEQQSGGAL IASTTSGLTI    600
EGTNSYGDFA YIRNSEAKNV VLENAGSLTV VTGSRAVDTI INANGKMDVY GKDVGTVLNS    660
AGTQTIYASA TSDKANIKGG KQTVYGLATE ANIESGEQIV DGGSTEKTHI NGGTQTVQNY    720
55 GKAINTDIVS GLQQIMANGT AEGSIINGGS QIVNEGGLAE NSVLNDGGTL DVREKGSATG    780
IQQSSQALV ATTRATRV TG TRADGVAFSI EQGAANNILL ANGGVLTVES DTSSDKTQVN    840
TGGREIVKTK ATATGTTLTG GEQIVEGVAN ETTINDGGIQ TVSANGEAIK TTINEGGTLT    900
VNDNGKATDI VQNSGAALQT STANGIEISG THQYGTFSIS GNLATNMLLE NGGNLLVLG    960
TEARDSTVGK GGAMQNGQD SATKVNSGGQ YTLGRSKDEF QALARAEDLQ VAGGTAIVYA    1020
60 GTLADASVSG ATGSLSLMTP RDNVTPVKLE GAIRITDSAT LTIGNGVDTT LADLTAASRG    1080
SVWLNSNNSC AGTSNCEYRV NSLLLNDGNV YLSAQTAAPA TTNGIYNTLT TNELSGSGNF    1140
YLHTNVAGSR GDQLVVNNNA TGNFKIFVQD TGVSPQSDDA MTLVKTTGGD ASFSLGNTGG    1200
FVDLGTIEYV LKSDGNSNWN LTNDVKPNPD PNPNPNPND PDKPDPKPD PKPDPTPEPT    1260
PTPVPEKRIT PSTAAVLNMA ATLPLVFDAE LNSIRERLNI MKASPHNNNV WGATYNTRNN    1320
65 VTTDAGAGFE QTLTGMTVGI DSPNDIPEGI ATLGAFMGYS HSHIGFDRGG HGSVGSYSLG    1380
GYASWEHESG FYLDGVVKLN RFESNVAGKM SSGGAANGSY HSNGLGGHIE TGMRTDGNW    1440
NLTPYASLTG FTADNPEYHL SNGMESKSVD TRSIYRELGA TLSYNMRLGN GMEIEPWLKA    1500
```

AVRKEFVDDN RVKVNNDGNF VNDLSGRRGI YQAGIKASFS STLSGHLGVG YSHGAGVESP 1560
WNAVAGVNWS F 1571
<212> Type : PRT
<211> Length : 1571
5 SequenceName : SEQ ID 2
 SequenceDescription :

Sequence

10 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
MAVKISGVLK DGTGKPVENC TIQLKARRNS ATVVVNTVAS ENPDEAGRYS MDVEYGQYSV 60
ILLVEGFPPS HAGTITVYED SQPGTLNDFL GAMTEDDVRP EALRRFELMV EEVARNASAV 120
AQNTAAAPKS ASDACTGARE AATHATDAAL SPANCTTSAG QAASSAQAS SSAGTASTKA 180
15 TEASKSAAAA ESSKSAAATS AGAAKTSETN AAVSQQSAAT SASTATTKAS EAASSARDAS 240
ASKEAAKSSE TSAASSASSA ASSATAAGNS AKAAKTSETN AKSSETAAEQ SASAAAGSKT 300
AAALSASAAS TSAGQASASA TAAGKSAESA ASSASTATTK AGEATEQASA AASSASAAKT 360
SETNAKASET SAESSKTAAS SSASSAASSA SSASASKDEA TRQASAAKSS ATTASTKATE 420
AAGSATAAAQ SKSTAESAAT RAETAAKRAE DIASAVALED ASTTKKGIVQ LSSATNSTSE 480
20 SLAATPKAVK AAYELANGKY TAQDATTQAK GIVQLSNATN STSEMLAATP KSVKAAYDLA 540
NGKYTAQDAT TAQKGIVQLS SATNSASETL AATPKAVKAA NDNANGRVPS ARKVNGKALS 600
SDITLTPKDI GTLNSTTMSF SGGAGWFKLA TVTMPQASSV VSITLIGGAG FNVGSPQQAG 660
ISELVLRAGN GNPKGITGAL WQRTSTGFTN FAWVNTSGDT YDIYVAIGNY ATGVNIQWDY 720
TSNASVTIHT SPAYSANKPE GLTDGTVYSL YTPSEQFYPP GAPIPWPSDT VPSGYALMQG 780
25 QTFDKSAYPK LAAAYPSGVI PDMRGWTIKG KPASGRAVLS QEQDGIKSHS HSASASSTDL 840
GTKTTSFSDY GTKSTNNTGA HTHSVSGTAA SAGNHTHSVT GASAVSQWSQ NGSVHKVSA 900
ASVNTSAAGA HTHSVSGTAA SAGAHAHTVG IGAHTHSVAI GSHGHTITVN AAGNAENTVK 960
NIAFNYIVRL A 971
<212> Type : PRT
30 <211> Length : 971
 SequenceName : SEQ ID 3
 SequenceDescription :

Sequence

35 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
MKRVITLFAV LLMGWSVNAW SFACKTANGT AIPIGGGSAN VYVNLAPAVN VGQNLVVDLS 60
TQIFCHNDYP ETITDYVTLQ RGSAYGGVLS NFSGTVKYSY SSYPFPTTSE TPRVVYNSRT 120
40 DKPWPVALYL TPVSSAGGVA IKAGSLIAVL ILRQTKYNYS DDFQFVWNIY ANNDVVVPTG 180
GCDVSARDVT VTLPDYPGSV PIPLTVYCAK SQNLGYLBSG TTADAGNSIF TNTASFSPAQ 240
GVGVQLTRNG TIIPANNTVS LGAVGTSASV LGLTANYART GGQVTAGNVQ SIIGVTFVYQ 300

<212> Type : PRT
45 <211> Length : 300
 SequenceName : SEQ ID 4
 SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
MGYVTGGLPM KNNRAWALIS GLILFSGTAP AADNLHFTGN LLGKSCTPVI NGNLLAEIHF 60
PTIAASDLMQ RGQSDRVPLV FQLKDCKSTT AFNVKVTLMG TEDTDLPGLF SIDSSSSATG 120
55 VGIGIETAGG AAVPINSTTG ASFPLNQGN SVNFNAWLQT VNGRNVTSGD FTATMTVTFE 180
YF 182
<212> Type : PRT
<211> Length : 182
 SequenceName : SEQ ID 5
60 SequenceDescription :

Sequence

65 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAVAVAL SLAAVTSVPA LAADKVVQAG 60
ETVNDGTLTN HDNQIVFGTA NGMTISTGLE LGPDSEENTG GQWIQNGGIA GNTTVITNGR 120

QVVLEGGTAS DTVIRDGGGQ SLNGLAVNTT LNNRGEQWVH EGGVATGTII NRDGYQSVKS 180
GGLATGTIIN TGAEGGPDS NSYTGQKVQG TAESTTINKN GRQIILFSGL ARDTLIYAGG 240
DQSVHGRALN TTLNGGYQYV HRDGLALNTV INEGGWQVVK AGGAAGNTTI NQNGELRVHA 300
GGEATAVTQN TGGALVTSTA ATVIGTNRLG NFTVENKAD GVVLESGGRL DVLESHSAQN 360
5 TLVDDGGTLA VSAGGKATSV TITSGGALIA DSGATVEGTN ASGKFSIDGT SGQASGLLLE 420
NGGSFTVNAG GQAGNTTVGH RGTTLAAGG SLSGRTQLSK GASMLNGDV VSTGDIVNAG 480
EIRFDNQTP NAALSRAVAK SNSPVTFHKL TTTNLTGQGG TINMRVRDLG SNASDQLVIN 540
GGQATGKTWL AFTNVGNSNL GVATTGQGIT VVDAQNGATT EEGAFALSRP LQAGAFNYTL 600
NRDSDEDWYL RSENAYRAEV PLYTSMLTQA MDYDRILAGS RSHQTGVNGE NNSVRLSIQG 660
10 GHLGHDNNGG IARGATPESS GSYGFVRLEG DLLRTEVAGM SLTTGVYGAA GHSSVDVKDD 720
DGSRAGTVRD DAGSLGGYLN LVHTSSGLWA DIVAQGTRHS MKASSDNNDP RARGWGWLGS 780
LETGLPFSIT DNLMLPQLQ YTWQGLSLDD GQDNAGYVKF GHGSAQHVRA GFRLGSHNDM 840
TFEGGTSSRD TLRDSAKHSV SELPVNWWVQ PSVIRTFSSR GDMSMGTAAG GSNMTFSPSR 900
NGTSLDLQAC LEARIRENET LEVQAGYAHV VEGESAEGYN GQATLNMTF 960
15 <212> Type : PRT
<211> Length : 949
SequenceName : SEQ ID 6
SequenceDescription :
20 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 MKKWHYIFCI ILFHLGLPCG YAANDGTCAT RGGTHTLSLN FPLTTVSAAN NVPGNTLIDI 60
ANATSSENYS VLCNCDSKHS NGAYHEIYYT ADPAPGMVYS TTASGLAFYY LNEYVDVGTK 120
ISVLNAGYTA VPFEHVSQA TTTDHTCQGN KTTAVGVSLK TGADAKISFR IKRSINGTVV 180
IPITDIALY ANISSTTTRG EAIKVRISG SLTAPQSCQI NAGQVIYFDF DTIPASEFSS 240
TAGQAITSRK ITKTVSIECT GMYERTQKV DASFTGTNRS SDDTMVATDN ADVGIKIYNK 300
SNAEVSNNNG KLPADMGNTT IFGRKNGSVT FSAAPASFTG ARPQPGVFNA TATLTIEFVN 360
30 <212> Type : PRT
<211> Length : 360
SequenceName : SEQ ID 7
SequenceDescription :
35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MSRYKTGHKQ PRFRYSVLAR CVAWANISVQ VLFPLAVTFT PVMAARAQHA VQPRLSMGNT 60
TVTADNNVEK NVAFAANAG TFLSSQPDSD ATRNFITGMA TAKANQEIQE WLKYGTVARV 120
KLNVDKDFSL KDSSLEMLYP IYDTPTNMLF TQGAHRTDD RTQSNIGFGW RHFSGNDWMA 180
GVNTFIDHDL SRSHTRIGVG AEYWRDYLKL SANGYIRASG WKKSPDIEDY QERPANGWDI 240
RAEGYLPAPW QLGASLMEYQ YGDEVGLFG KDKRQKDPHA ISAEVYTPV PLLTSLAGHK 300
45 QGKSGENDTR FGLEVNYRIG EPLAKQLDTR SIRERRVLG SRYDLVERN NIVLEYRKSE 360
VIRIALPERI EGKGGQTLST GLVVSKATHG LKNVQWEAPS LLAEGGKITG QGSQWQVTL 420
AYRPGKDNYY AISAVAYDNK GNTSKRVQTE VVITGAGMSA DRTALTLDGQ SRIQMLANGN 480
EQKPLVLSLR DAEGQPVGTG KDQIKTELTF KPAGNIVTRS LKATKSQAKP TLGEFTETEA 540
GVYQSVFTTG TQSGEATITV SVDGMSKTVT AELRATMMDV ANSTLSANEP SGDVVADGQQ 600
50 AYTLTLTAVD SEGPNVTGEA SRLRFVPQDT NGVTVGASE IKPGVYSAV SSTRAGNVVV 660
RAFSEQYQLG TLQQTLLKFA GPLDAAHSSI TLNPDKPVG GTVTAIWTVK DAYDNPVTS 720
TPEAPSLAGA AAEGSTASGW TNNGDGTWTA QITLGSTAGE LEVMPKLNGQ NAAANAAKVT 780
VVADALSSNQ SKVSVAEDHV KAGESTTVTL VAKDAHGNAI SGLALSASLT GTASEGATVS 840
SWTEKNGSY VATLTTGGKT GELRVMLPFN GQPAATEAAQ LTVIAGEMSS ANSTLVADNK 900
55 APTVKTTTEL TFTVKDAYGN PVTGLKPDAP VFSGAAGTGS ERPSAGNWE KNGVYVSTL 960
TLGSAAGQLS VMPRVNGQNA VAQPLVLNVA GDASKAEIRD MTVKVNNQLA NGQSANQITL 1020
TVVDTYGNPL QGQEVTLTLP QGVTSKTGNT VTTNAAGKAD IELMSTVAGE HNISASVNGA 1080
QKTVTVKFNA DASTGQANLQ VDAAAQKVN GKDAFTLTAN VEDKNGNPVP GSVTFNLPR 1140
GVKPLTGDNV WVKANDEGKA ELQVVSVTAG TYEITASAGN SQPSNTQTIT FVADKATATV 1200
60 SGIEVIGNYA LADGNAKQTY KVTVTDANN LLKDSEVTLT ASPANLVLT NGTAKTNEQG 1260
QAIFTATTTV AAKYTLTAKV SQADGQESTK TAESKFVADD TNAVLTASSD VTSLVADGIS 1320
TAKLEVTLMS ANNPVGGNMW VDIKTPEGVT EKDYQFLPSK NDHFVSGKIT RTFSTSKPGV 1380
YTFTFNALTY GGYEMKPVTV TITAVDADTA KGEEAMN 1417
<212> Type : PRT
65 <211> Length : 1417
SequenceName : SEQ ID 8
SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MARGWASSEA SGAMTDWLN FGTARISLGV DEDFSLKNSQ FDFLHPWYDT PDYLLFSQHT 60
LHRTDDRTQI NTGLGWRHFT SSWMSGINLF FDHDLTRYHS RAGLGAEYWR DYLLKLSSNAY 120
IGLTGWSRAP ELDNDFEARP ANGWDLRAEG WLPAPWQLGG KLVYEQYYGD EVALFDKNDR 180
QSNPHAITAG LNYTPFPLLT LSAEQRQGKQ GENDTRFAVD LTWQPSSSMQ KQLNPDEVAG 240
10 RRSLAGSRYD LIDRNNNIVL EYRKELIRL SLLDPVKGKS GEIKPLVSSL QTKYALKGYN 300
IEAAALEAAG GKVSTSGKDI TVTLPGYRFT NTPETDNTWS IDVTAEDVKG NLSRHEQSMV 360
VIQAPTLSQK DSSLNVNPLT VAADKKSTTT LTVTAHDSGD TPVPGALALQT RSEGVQDITL 420
SDWTDNGDGS YTQILTAGTT SGSVTLTPQI NGESAVKESI VVNIVPVVSS RDHSSITIDN 480
VSYYAGDDIK VRVELKDDSN QPVAYQKEEL VKAVTVENSK PGATIVWHEE QPGVYAANYF 540
15 AYKQGTALRA QLSLHNWNA LQSHIYNIEA NQNKARVATL SATNNDVYAD KKTFTNTLTIN 600
VTDESNDPLT NHQVTFKNEK GSAEFVEPPQ QNTDAYGVAT INMVSQVAEE NTISATLPNG 660
FSQRIIAKFV SDSSTPKFKQ LVADPDITIA GNSQGSTLTA IITDFHNNPL KDMKVNFBVAP 720
GGSQLDNTTA TTDQSGIVRV HLTSSKAGSY SVDASLEVSK NIHQSVTITV VPNREQSVMT 780
LNAGSGSAIA NNTNIVTLTA SVKDVYGHPL PDEDVKFTLP ASMTGNFTLS SETARTDANG 840
20 DAVVTLRGTK AGEFTVTATL TRNNTVAYQQ VTFIGDTNSA QLQPLTASLN SIVAGNSTGS 900
TLTATILDAY QNPLKDQLVT FQSDNVTLSL TEVTTNTLGG ATVTMTSNIA GQHNVVVSRK 960
AQASDNKTFS LSVLPDESSA KVISITGAEK TITVGENITL RILVQDAFNN VIAGQVRVLS 1020
AQPTTNITIG DTAYTDNNGY AYVNLSTQP GVIQVATATL NNSSSKVDVN VANGKLELTS 1080
SKPETTVHNS EGITLTATAR NARGELMPGQ IITFSVTPEG ATLSNTGEVL TDQSGQAKVT 1140
25 LTSDKVNVTY VTAIMGKDVV VQSQVTVAVK ADAKTAHVVS VVASPDITTA DGIDSSTITS 1200
RVEDDYGFVP EGVDISHGLD TKGSPVVNIP TTRTDQSGQV TATITSTLAE TLTNVVQVPG 1260
TANQSATITL VAGTADESKS ILKSDVDTLK ADYQQSAKLT LTLQDKYGNP IVTSDHLEFV 1320
QSGPFVNFLK LSDIDYSQRN YGEYTVTVTG GKEGTATLIP MLNGVHQANL SISLNLIQSI 1380
KEMSGHVTAN NHTFSTAKFP SEGFAAYYT LNNDNFEGK TVDDYMFSSS QGWVSVDASG 1440
30 KVSFANIGDQ TSVTISAVPR QGGTTYQTLI KLKGWVWNG NHTNIWLAAN ALCHAKNDGY 1500
NLPGITHLTS GENKRTQGS LYGWGNVGAF SSNSQFTPGA YWTSESDDYS RHYYVQMLTG 1560
MTGSDADSSP QLTACRSL 1579
<212> Type : PRT
<211> Length : 1579
35 SequenceName : SEQ ID 9
SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MITHGCTRT RHKKHLKKT LMLSAGLGLF FYVNQNSFAN GENYFKLGSD SKLLTHDSYQ 60
NRLFYTLKTG ETVADLSKSQ DINLSTIWSL NKHLYSSESE MMKAAPGQOI ILPLKKLPFE 120
YSALPLLGS PLVAAGGVAG HTNKLTKMSP DVTKSNMTDD KALNYAAQQA ASLGSQLOSR 180
45 SLNGDYAKDT ALGIAGNQAS SQLQAWLQHY GTAEVNLQSG NNFDSGLDF LLPFYDSEKM 240
LAFGQVGARY IDSRFTANLG AGQRFPLPAN MLGYNVFIQ DFSGDNTRLG IGGEYWRDYF 300
KSSVNGYFRM SGWHESYNKK DYDERPANGF DIRFNGYLP YPALGAKLIY EQYYGDNVAL 360
FNSDKLQSNP GAATVGVNYT PIPLVTMGID YRHGTGNEND LLYSMQFRYQ FDKSWSQQIE 420
PQYVNELR TLSGSRYDLVQR NNNIILEYKK QDILSLNIPH DINGTEHSTQ KIQLIVKSKY 480
50 GLDRIVWDDS ALRSQGGQIQ HSGSQSAQDY QAILPAYVQG GSNIYKVTAR AYDRNGNSSN 540
NVQLTITVLS NGQVVDQGV TDFTADKTSA KADNADTITY TATVKKNGVA QANVPVSFNI 600
VSGTATLGAN SAKTDANGKA TVTLKSSTPG QVVVSAKTAE MTSALNASAV IFFDQTKASI 660
TEIKADKTTA VANGKDAIKY TVKVMKNGQP VNNQSVTFST NFGMFNGKSQ TQATTGNDGR 720
ATITLTSSSA GKATVSATVS DGAEVKATEV TFFDELKIDN KVDIIGNNVR GELPNLWLQY 780
55 GQFKLKASGG DGTYSWYSEN TSIATVDASG KVTLNGKGSV VIKATSGDKQ TVSYTIKAPS 840
YMIKVDKQAY YADAMSICKN LLPSTQTVLS DIYDSWGAAN KYSHYSSMNS ITAWIKQTSS 900
EQRSGVSSTY NLITQNPLPG VNVNTPNVYA VCVE 934
<212> Type : PRT
<211> Length : 934
60 SequenceName : SEQ ID 10
SequenceDescription :

Sequence

65 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MLVLSESEFKN KLLPMNGYMK GGSDSGSKAQ ARATEKGIEL QREMWQTNMQ NLAPFTPLAQ 60

QYVSQQLQNL SLQGQGGQALN QYYNSQOYKD LAGQARYQSL AAAEATGGLG STATGNQLAA 120
IAPTLGQNLW SGQMNNYNNL ANIGLGALTG QANAGQNYAN NVSQLYQQQA AASAANANKP 180
SGLQSFATGA IGGAASGAMI GSAVPVIGTG IGALAGGVIG GLGSLF 226
<212> Type : PRT
5 <211> Length : 226
SequenceName : SEQ ID 11
SequenceDescription :

Sequence
10 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKILSGLIL LLCCPYGFAA NGDGATHMSN LSFGLTVAA ANNHSGYNIF EALSNTTGTY 60
PVRCHCDDTH GGPGQQTAF FIFYTGDAAAP GLVLERTLNG LNYVALNDYL SVGVTFIFIIN 120
15 NQYAAIPFEH LSNQSTSPQH TCGAGNNGST VNLDSEGRSAK LSFYVRHSIT GTVTIPTTEV 180
AWLYAGMSDH FPKTTPVSKV TIRGQLTAPQ NCELTPNQSI DVDFQKINSA EFSSTAGSII 240
AERKIKTEVT VSCTGMEDVR STEVVSASMI AANRSADATM IVTSNPDVGI KIFDKNDRPV 300
NVDGGNLPAD MGAISRLGKT DGSVTFYSAP ASLTGAKPAP DNGFTATATL VIEFTN 356

20 <212> Type : PRT
<211> Length : 356
SequenceName : SEQ ID 12
SequenceDescription :

25 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNKIYRLKWN RSRNCWSVCS ELGSRVKGKK SRAVLISAIS LYSSLVFADD VIVNQDKTID 60
30 F'GKENQSIDY RITVTDNANL VINATDTSRP RLTLASGGGL DITGGKVTIN GPLNFKLLKGT 120
GFLNVSNAAGS ELYADDLYES NSGMRHDRGY FNVSNGGKIH VKGTSRLTYL QGNVSGEGSQ 180
VNSETFFMGV YGSYGGNQYL SVNNGGEVNA RKQISLGYD QVSDTTLAVS EGGKISAPTI 240
SLSTNSELAL GAQEGSAAKA AGIIDAIEKIE FVWAKTSEKK ITLNHTDKDA TISADIVSGS 300
EGLGYINALN GTTYLTGDNS AFSGKVKIEQ NGALGITQNI GTAEINNNGK LHLKADDSMT 360
35 FANKISGNGT ISIDSGTVEL TGNNYAFSGY IDVASGAVAV ISEDKNIGRA ELDVDGKLQI 420
NANKDWVFDN DLEGRGIVEI NMGNHEFSFD EFAYTDWFQG SLAFQNTTFN LEKNAEFLQK 480
GGITAGQGS LVTVGKAHSI STLGFSGGTV DFGALTAGAQ MTEGTVNVSK TDLRGEVVI 540
QVSDSDVVR VSRDIDSALS LTEVDDGNST IKLVDAQGAE VLGDAGNLQL QDKNGQILSS 600
SAQRDIQONG QKAAVGTIDY RLTSVGNNDG LYIGYGLTQL DLHATDSAL VLSSNGKSEN 660
40 AADLSAKITG SGDLAFSSQK GQTVSLSNKD NDYTGVTDLR SGTLLLNNDN VLGNTHLRL 720
AAETELDMNG HSQTVGTLNG SADSLSLNG GSLTVTNGGT STGSLTGSSE LNIQGGTLDI 780
AGDNSNLNAN VNIANSANVL VSHAQGLGSA NVENNGTLAL NNSAEKRAAA SVNYALGGNL 840
TNNGTLMTGM SGQQAGNVLV VKGNYHGNG QLVMTVLNG DDSVTDKLVV EGDTSCTTAV 900
TVNNAGGTGA KTLNGIELIH VDGKSEGEFV QAGRIVAGAY DYTLAGQGA NSGNWYLTSG 960
45 SDSPQLQPEP DPMPNPEPNP NPEPNPNPTP TPGPDNLVDN DLRPEAGSYI ANLAAANTMF 1020
TTRLHERLGN TYYTDMVTGE QKQTTMWMRH EGGHNKWRDG SGQLKTQSNR YVLQLGGDVA 1080
QWSQNGSDRW HVGVMAGYGN SDSKTISSRT GYRAKASVNG YSTGLYATWY ADDESNGAY 1140
LDSWAQYSWF DNTVKGDDLQ SESYKSKGFT ASLEAGYKHK LAEFNGSQGT RNEWYVQPQA 1200
QVTWMGVKAD KHRESNGTLV HSNLGDGNVQT RLGVKTWLKS HHKMDDGKSR EFQPFVEVNW 1260
50 LHNSKDFSTS MDGVSVTQDG ARNIAEIKTG VEGQLNANLN VWGNVGVQVA DRGYNDTSAM 1320
VGIKWQF 1327
<212> Type : PRT
<211> Length : 1327
SequenceName : SEQ ID 13
55 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
60 <400> PreSequenceString :
MITMKKSVLT AFITVVCATS SVMAADDNAI TDGSVTFNGK VIAPACTLVA ATKDSVVTLP 60
DVSATKLQTN GQVSGVQTDV PIELKDCDTT VTKNATFTFN GTADTTQITA FANQASSDAA 120
TNVALQMYMN DGTTAIKPDT ETGNILLQDG DQTLTFKVDY IATGKATSGN VNAVTFNHIN 180
YY 182
65 <212> Type : PRT
<211> Length : 182
SequenceName : SEQ ID 14

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSKFVKTAIA ATMVMGAFAS TSTIAAGNNG TARFYGTIED SPCSIVPDDH KLEVDMGDIG 60
SGILKNNGTS TPKAFQIHLQ DCVFDTQTTM TTTFTGNASS TNSGNYTIIY NTDTGAAFNN 120
VSLAIGDAQG TSYKSGAGIE QKIVNDTATN KGKAKQTLDF KAWLVGAADA PDLGNFEANT 180
10 TFQITYL 187
<212> Type : PRT
<211> Length : 187
SequenceName : SEQ ID 15
SequenceDescription :

Sequence

20 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRVIFLRKEY LSLPLSMIAS LFSANGVAAA IDLCQGYDIK ASCHASRQSL SGITQVWSIA 60
DGQWLVSFSDM TNNASGGAVF LQQGAEFTLS PENETGMTLF ANNTVSGEYN NGGAIFAKEN 120
STLNLTDVIF SGNVAGGYGG AIYSSGTNDT GAIDLRVTNA VFRNNIANDG KGGAIYTINN 180
DIYLSDDVFN NNQAYTSTSY SDGDGGGADV TDNNSDSKHP SGYTIINNTA FTNNTAEGYG 240
GAITYTNSATA PYLIDISVDD SYSQNGGVLV DENNSAAGYG DGPSSAAGGF MYLGLSEVTF 300
25 DIADGKTLVI GNTENDGAVD SIAGTGLITK TGSGDLVLNA DNNDFTGEMQ IENG EVT LGR 360
SNSLMNVGDT HCQDDPQDCY GLTIGSIDKY QNQAELNVGS TQOTFAHSLT GFQNGTLNID 420
AGGNVTVNQG SFAGTIEGAG QLTIAQNGSY VLAGAQSMAL TGDIVVDAGA VLSLEGDAAD 480
LAALQDDPQS IVLNGGMLDL SDFSTWQSGT SYKDGLEVSG SSGTVIGSQD VVDLAGGNDM 540
HIGGDGKDG VYVIDAGDGQ VSLANDNQYL GTTQIASGTL MVSDNSQLGY THYNRQVIFT 600
30 DKPQESVMEI TANVDTRSTT TEHGRDIEMR ADGEVAVDAG VDTQWGALMA DSSGQHQDEG 660
STLTKTGAGT LELTASGTTQ SAVRVEEGLT QGDVADIFPY ASSLWVG DGA TFVTGADQDI 720
QSIDATSSGT IDISDGTVLR LTGQDTSVAL NASLFNC DGT LVNATDGVTL TGE LNTNLET 780
DSLTYLSNVT VNGNLNTSG AVSLQNGVAG DTLTVNGDYT GGGTLLLDSE LNGDDSVSDQ 840
LVMNGNTAGN TTVVNSITG IGEPTSTGIK VVDFAADPTQ FQNN AQFSLA GSGYVNMGAY 900
35 DYTTLVEDNND WYLRSEQVTP PSPDPDPPTP DPDP TQDPDP TPDPEPTPAY QPVLNAKVGG 960
YLNNLRAANQ AFMMERRDHA GGDGQTLNLR VIGGDYHYTA AGQLAQHEDT STVQLSGDLF 1020
SGRWGTDGEW MLGIVGGYSD NQGDSRSSMT GTRADNQNHG YAVGLTSSWF QHKGKQKQAW 1080
LDNWLQYAWF SNDVSEHEDG VDHYHSSGII ASLEAGYQWL PGRGVVIEPQ AQVIYQGVQQ 1140
DDFTAANRAR VSQSQGGDIQ TRLGLHSEWR TAVHVIPTLD LNYHYDHPST EIEEDASTIS 1200
40 DDAVKQRGEI KVGVTGNISQ RVSLRGSVAW QKGSDDFAQT AGFLSMTVKW 1250
<212> Type : PRT
<211> Length : 1250
SequenceName : SEQ ID 16
SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MHSWKKKLIV SQLALACTLA ITSQANAATN DISGQTYNTF HHYN DATYAD DVYYDGYVGW 60
NNYAADSYYN GDIYPVINNA TVNGVISTYY LDDGISTNTN ANSLTIKNST IHGMITSECM 120
TTDCADDRAT GYVYDRLTLS VDNSTIDDNY EHYTYNGTYN NAADTHVVDV YDMGTAITLD 180
QEV DLSITNN SHVAGITLTQ GYEWEDIDDN TVSTGVNSSE VFNN TITVKD STVTSGSWTD 240
EGTTGWFGHT GNASNYSNTL TADDVAIAAI ANPYADNAMQ TTVTLDNSTL MGDVVFSSNF 300
55 DENFFPQGAN SYRDADGDVD TNGWDGTD RM DVT LNNGSKW VGAAMSVH MV DEDGDGSYDG 360
YAVGTEATAT LLDIAANSLW PSSTVGVDNI NTQYDENGHI VGNEVYQSGL FNVTLNGGSE 420
WDTTKSSLID TLSINSGSQV NVADSR LID TVSLTGGSNL NIGEDGHVAT NTLTIDNSTV 480
KMSDDVSAGW GLEDAALYAN TITVTNDGLL DINVDQFDAN PFQADTLNLT STTD TNGNIH 540
AGVFDIHSSD YVMDTDLVND RTNDTTKSNY GYGLIAMNSD GHLTINGNGD NDNTASIEAG 600
60 QNEVDNNGDH VAAATGNYKV RIDNATGAGS IADYNGNELI YVNDKNSNAT FSAANKADLG 660
AYTYQAEQRG NTVVLQQMEL TDYANMALSI PSANTNIWNL EQDTVGT RL T NSRHGLADNG 720
GAWVS YFGGN FNGDN GTINY DQDVNGIMVG VDTKIDG NNA KWIVGAAAGF AKGDMNDRSG 780
QVDQDSQTAY IYSSAHFANN VFVDGSLSYS HFNNDLSATM SNGTYVDGST NSDAWGFG LK 840
AGYDFKL GDA GYVTPYGSIS GLFQSGDDYQ LSNDMKVDGQ SYDSMRYELG VDAGYTFTYS 900
65 EDQALTPYFK LAYVYDDSN DNDVNGDSID NGTEGSAVRV GLGTQFSFTK NFSAYTDANY 960
LGGGDVDQDW SANVGKYTW 980
<212> Type : PRT

<211> Length : 980
SequenceName : SEQ ID 17
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
10 MKLKHVGMIV VSVLAMSSAA VSAAEGDESV TTTVNGGVIH FKGEVVNAAC AIDSESMNQT 60
VELGQVRSSR LAKAGDLSSA VGFNIKLNDC DTNVSSNAAV AFLGTTVTSN DDTLALQSSA 120
AGSAQNVGIQ ILDRTGEVLI LDGATFSAKT DLIDGTNILP FQARYIALGQ SVAGTANADA 180
TFKVQYL 187
<212> Type : PRT
<211> Length : 187
15 SequenceName : SEQ ID 18
SequenceDescription :

Sequence

20 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKLLKVAALIA AIVFSGSALA GVVPQYGGGG GNHGGGGNNS GPNSELNIYQ YGGGNSALAL 60
QADARNSDLT ITQHGGGNGA DVGQGSDDSS IDLTQRGFGN SATLDQWNGK DSHMTVKQFG 120
GGNGAAVDQT ASNSTVNVVTQ VGFGNNATAH QY 152
25 <212> Type : PRT
<211> Length : 152
SequenceName : SEQ ID 19
SequenceDescription :

30 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 MPIGNLGHNP NVNNSIPPAP PLPSQTDGAG GRGQLINSTG PLGSRALFTP VRNSMADSGD 60
NRASDVPGLP VNPMRLAASE ITLNDGFEVL HDHGPLDTLN RQIGSSVFRV ETQEDGKHIA 120
VGQRNGVETS VVLSDQEYAR LQSIDPEGKD KVFVTGGRGG AGHAMVTVAS DITEARQRIL 180
ELLEPKGTGE SKGAGESKGV GELRESNSGA ENTTETQTST STSSLRSDPK LWLALGTVA 240
GLIGLAATGI VQALALTPEP DSPTTTDPDA AASATETATR DQLTKEAFQN PDNQKVNIDE 300
LGNAIPSGVL KDDVVANIEE QAKAAGEEAK QQAIENNAQA QKKYDEQQAK RQEELKVSSG 360
40 AGYGLSGALI LGGGIGVAVT AALHRKNQPV EQTTTTTTTT TTTSARTVEN KPANNTPAQG 420
NVDTPGSEDT MESRRSSMAS TSSTFFDTS IGTVQNPYAD VKTSLHDSQV PTSNSNTSVQ 480
NMGNTDSVVY STIQHPPRDT TDNGARLLGN PSAGIQSTYA RLALSGGLRH DMGGLTGGSN 540
SAVNTSNNPP APGSHRFV 558
<212> Type : PRT
45 <211> Length : 558
SequenceName : SEQ ID 20
SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MFSTFKKAAL LAALALPFST MAAPT VTFQG EVTDQTCSVN INGQTNVVL MPTVAMADFG 60
ATLADGQSAG QTPFTVSVSN CQAPT GADQA INTTFLGYDV DASTGVMGNR DTSSDAAKGF 120
55 GIQLMDSSTS GNPVTLGAT NVPGLTLKVG DTEASYDFGA RYFVIDSAAA TAGKITAVAE 180
YTL SYL 186
<212> Type : PRT
<211> Length : 186
SequenceName : SEQ ID 21
60 SequenceDescription :

Sequence

65 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNSEGGKPGN VLTVNGNYTG NNGLMTFNAT LGGDNSTPTDK MNVKGDTQGN TRVRVDNIGG 60
VGAQTVNGIE LIEVGGNSAG NFALTGTGVE AGAYVYTLAK GKGNDKKNWY LTSKWDGVTP 120

ADTPDPINNP PVVDPEGPSV YRPEAGSYIS NIAAANSIFS HRLHDRLGEP QYTDSLHSQD 180
SASSMWMRHV GGHERRSAGD GQLNTQANRY VLQLGGDLAQ WSSNAQDRWH LGVMAGYANQ 240
HSNTQSNRVG YKSDGRISGY SAGLYATWYQ NDANKTGAYV DSWALYNWFD NSVSSDNRSA 300
DDYDSRGVTA SVEGGYTFEA GTCGSGSEGL NTWYVQPQAO ITWMGVKDSH HARKDGTRIE 360
5 TEGDGNVQTR LGVKTYLNSH HQRDDGKQRE FQPYIEANWI NNSKVYAVKM NGQTVSRDGA 420
RNLGEVVRTGV EAKVNNNLNL WGNVGVQLGD KGYSQTQGML GVKYSW 466
<212> Type : PRT
<211> Length : 466
SequenceName : SEQ ID 22
10 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
15 <400> PreSequenceString :
MSYLNLRLYQ RNTQCLHIRK HRLAGFFVRL FVACAFVQA PLSSAELYFN PRFLADDPQA 60
VADLSRFENG QELPPGTYRV DIYLNNGYMA TRDVTFTNGD SEQGIVPCLT RAQLASMGLN 120
TASVAGMNL ADDACVPLTT MVQDATAHLD VGQQLNLTI PQAFMSNRAR GYIPPELWDP 180
GINAGLLNIN FSGNSVQNRG GGNSHYAYLN LQSGNLIGAW RLRDNTTWSY NSSDRSSGSK 240
20 NKWQHINTWL ERDIIPLRSR LTLGDGYTQG DIFDGINFRG AQLASDDNML PDSQRGFAPV 300
IHGIARGTAQ VTIKQNGYDI YNSTVPPGPF TINDIYAAGN SGDLQVTIKE ADGSTQIFTV 360
PYSSVPLLQR EGHTRYSTA GEYRSGNAQQ EKPRFFQSTL LHGLPAGWTI YGGTQLADRY 420
RAFNFGIGKN MGALGALSVD MTQANSTLPD DSQHDGQSVR FLYNKSLNES GTNIQLVGYR 480
YSTSGYFNFA DTTYSRMNGY NIETQDGVIV VKPKFTDYNN LAYNKRGLQ LTVTQQLGRS 540
25 STLYLSGSHQ TYWGTSNVDE QFQAGLNTAF EDINWTLNYS LTKNAWQKGR DQMLARNVNI 600
PFSHWLRSDS KSQWRHASAS YSMHDLNLR MTNLAGVYGT LLEDNNLSYS VQTGYAGGGD 660
GNSGSTGYAT LNYRGGYGNA NIGYSHSDDI KQLYYGVSGG VLAHANGVTL GQPLNDTVVL 720
VKAPGAKDAK VENQTGVRTD WRGYAVLPYA TEYRENVAL DTNTLADNVD LDNAVAVNVP 780
TRGAIVRAEF KARVGKLLM TLTHNNKPLP FGAMVTSESS QSSGIVADNG QVYLSGMPLA 840
30 GKVVQVKGEE ENAHCVANYQ LPPESQQQLL TQLSAECR 878
<212> Type : PRT
<211> Length : 878
SequenceName : SEQ ID 23
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MQIIFGEKCV SLLRLFFAAV LMLWCAQTAA YSGQCHTTQG NPYIGVNFV KTLEEEENTT 60
GVVKDKFYQW NESNDYYVSC DCDKDNVRSR RWAFADSPV VYLGDNWYKI NDYLAQVLL 120
QVKGSSPTAV PFENVGTGAD TRWHICDPGG QRLGGQGASG NSGSFSLKIL QPFVGSVIP 180
PMALARLFEC YNIPAGDSCT TTGTPVLVYY LSGTINSLGS CSVNAGETIE VDLGDFVFAAN 240
FRVVGHKPLG ARTAELAIPV RCNTGNAGLV NVNLSLTATT DPSYPQAIKT SRPGVGVVVT 300
45 DSQNNIISPA GGTLPPLSIPD DADSIA 326
<212> Type : PRT
<211> Length : 326
SequenceName : SEQ ID 24
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 MKIKTLAIVV LSALSLSSTA ALAAATTVNG GTVHFKGEVV NAACAVDAGS VDQTVQLGQV 60
RTASLAQDGA TSSAVGFNIQ LNDCTNVAS KAAVAFGLTV IDAGHTNVLA LQSSAAGSAT 120
NVGVQILDRT GAALTLDGAT FSEQTTLNNG TNTIPFQARY YAIGEATPGA ANADATFKVQ 180
YQ 182
<212> Type : PRT
60 <211> Length : 182
SequenceName : SEQ ID 25
SequenceDescription :

65 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

MKLVKVIATLI ATVAVGVSN SNFASASTTS ASLTVNSNLT MGTCSAQIMD NSNKVINEVV 60
FGNVYISELG AKSKVQQFKI RFSNCSGLPQ NSAQIVLAPN GISCAGSQSS SAGFSNKFTD 120
ASAATRTAVE VWTDTTPESN GSTQFHCAQK IPVPVTLPAD TTTQPYDYPL SARMTVAEGR 180
LVTDVRPGNF RSPTTFTITY Q 201

5 <212> Type : PRT
<211> Length : 201
SequenceName : SEQ ID 26
SequenceDescription :

10 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

15 MASTVEYGET VDGVVLEKDI QLVYGTANNT KINFGGEGHI KEFGISSNTE INGGYQYIEM 60
NGTAEYSVLN DGYQIVQMG G AANQTTLNNG VLQVYGAAND PTIKGGRLIV EKDGITVLAA 120
IEKGGLLEV K EGGLAIAVDQ KAGGAIKAST RVMEVFGTNR LGQFEIKNGI ANNMLLENGG 180
SLRVEENDFA YNTTVDSGGL LEVMDGGTAT GVDKAGGKL IVSTNALEVS GTNSKGQFSI 240
KDGVSKNYEL DDGSGLIVME DTQAIDTILD EHATMQSLGK DTGTRVQANA VYDLGRSDQN 300
GSITYSSKAI SENMVINNGR ANVWAGTMVN VSVRGNDGIL EVMKPQINYA PAMLVGKVVV 360
20 SEGASLRTHG AVDTSKADVS LENSATIIA DITTTNQNTR LNLANLAMSG ANVIMMDES V 420
TRSSVTASAE NFFTLLTNTL SGNGNFYMR T DMANHQS DQL NVTGQATGDF KIFVTD TGAS 480
PAAGDSLTLV TTGGGDAAFT LGNAGGVVDI GTYEYTL LDN GNHSWSLAEN RAQITPSTTD 540
VLNMAAAQPL VFDAELDTVR ERLGSVKGV S YDTAMWSSAI NTRNNVT TDA GAGFEQTLTG 600
LTLGIDSRFS REESSTIRGL FFGYSHSDIG FDRGGKGNVD SYTLGAYAGW EHONGAYVDG 660
25 VVKVDRFANT IHGKMSNGAT AFGDYN SNGA GAHVESGFRW VDGLWSVRPY LAFTGFTTDG 720
QDYTL SNGMR ADVGNTRILR AEAGTAVSYH MDLQNGTTLE PWLKA AVRQE YADSNQVKVN 780
DDGKFNDVA GTRGVYQAGI RSSFTPTLSG HLSVSYNGA GVESPWNTQA GVVWTF 836

<212> Type : PRT
30 <211> Length : 836
SequenceName : SEQ ID 27
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

40 MQRKGNKLLI QLCSVILLFF TTSWYALANE CYIERN AEGD YHMKISSTQL SLASQMVEVP 60
TEIAEATWDV NIQLRGDAIG CKSLGDSKAV HFLNTADPSL ISTYTTTNGA ALLKTTVPGI 120
VYSVELLCLS CGAADEL DLW LPAQSGADNF IPSTQTKWAY EYSDQSWYLR FRLFITPEFK 180
PKNGVSSGTT IAGKIASWYI GTNDQPWINF YIDNDSLKFF VDEPTCATVA LAQDQGNVSG 240
NQVTLGNSYV SEVKNGLTRE IPFSIRAEYC YASKITVKLK AANKPSDATL VGKTTGSASG 300
VAVKVNSTYD NSKVLLKADG SNTVDYNFAA WSNLLFLPF TAQLVPDGS G NAVGVGTFSG 360
NATFSFTYE 369

45 <212> Type : PRT
<211> Length : 369
SequenceName : SEQ ID 28
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

55 MYQFTHQKSR IPKKTLLAAC CALFYSSNGA AADTVEYDSS FLMGTGASTI DVKRYAQGNP 60
TPPGLYNVRV FVNGQATSSL EIPFVDIGEN SAAACLTHKN LAQLHIKQPE QPVTLLAREG 120
EEEDCLDLAK SYEKADVCFD GSDQFLDLTI PQAYVLKSYG GYVDPSLWES GINAATLAYT 180
LNAYHTSSDN DNSDSVYGAF NSGINLGAWH FRARGNYNWT TDNGSDFDFQ DRYLORDIPA 240
IRSQIIMGDA YTTGETFDSV NVRGVRLYSD SRMLPSALAS YAPTIRGVAN SNAKVTVTQS 300
GYKIYETTVP PGEFVIDDIS PSGFGSELV TIEEADGSKR TFTQPFSSV QMRPGVGRW 360
60 DFSAGKVIDD SLRSEPNMGQ ASYYYGLNNL FTGYTGIOFT DNNYLAGLLG VGINTSIGAF 420
AVDVTHSRAE IPDDKTYQGQ SYRV TWNKLF QDTGTSFNLA AYRYSTQDYL GLHDALVLID 480
DAKHLSADED KNTMQTYSRM KNQFTVSINQ PLNIAYEDYG SLFISGSWTY YWAANNSRTE 540
YNVGYSKSVS WGSF SVNLQR SWNEGEKDD AMYVSVSVPI ENILGGKRKS SGFRNLNTQL 600
NTDFDGS HQ L NVNSSGNTEN NLVNYSVNAG YSLDKNAGDL ASVGGYLYNE SGLGGISASA 660
65 SATSDNSQQY SISTDGGFVL HSGGLTFTNN SFSSNDTLVL INALGAKGAR INNSNNEIDR 720
WGYAVTSSVS PYREN RVGLN IETLENDVEL KSTSATTVPR SGSVLTRFE TDEGRSAVLN 780
ITAANGKSIP FAAEVYQGEV MIGSMGQGGQ AFVRGINDSG ELIVRWYENN QTIDCKLHYQ 840

FPAQPQTQGS TNTLLLNLT CQVANH 866
<212> Type : PRT
<211> Length : 866
SequenceName : SEQ ID 29
5 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
10 <400> PreSequenceString :
MKFKRLLHSG IASLSLVACG VNAATDLGPA GDIHFSITIT TKACEMEKSD LEVDMGMTMTL 60
QKPAAVGTVL SKKDFTEELK ECDGISKATV EMDSQSDSDS DSMFALEAGG ATGVALKIED 120
DKGTQQVPKG SSGTPIEWAI DGETTSLHYQ ASYVVVNTQA TGGTANALVN FSITYE 176

<212> Type : PRT
<211> Length : 176
SequenceName : SEQ ID 30
15 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
20 MKYNNIIFLG LCLGLTTYSA LSADSVIKIS GRVLDYGCTV SSDSLNFTVD LQKNSARQFP 60
25 TTGSTSPAVP FQITLSECSK GTTGVRVAFN GIEDAENNTL LKLDEGSNTA SGLGIEILDG 120
NMRPVKLNLDL HAGMQWIPLV PEQNNILPYS ARLKSTQKSV NPGLVRASAT FTLEFQ 176

<212> Type : PRT
<211> Length : 176
30 SequenceName : SEQ ID 31
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
35 <400> PreSequenceString :
MKWRKRGYLL AAILALASAT IQAADVTITV NGKVVAKPCT VSTTNATVDL GDLYSFSLMS 60
AGAASAWHDV ALELTNCPVG TSRVTASFSG AADSTGYKYN QGTAQNIQLE LQDDSGNTLN 120
TGATKTVQVD DSSQSAHFPL QVRALTVNGG ATQGTIQAVI SITYTYS 167

<212> Type : PRT
<211> Length : 167
SequenceName : SEQ ID 32
40 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
45 MKRAPLITGL LLISTSCAYA SSEGCGADST SGATNYSSVV DDVTVNQTDN VTGREFTSAT 60
50 LSSTNWQYAC SCSAGKAVKL VYMVSPVLT TGHQTGYKYL NDSLDIKTMN RPGNPGD 117

<212> Type : PRT
<211> Length : 117
SequenceName : SEQ ID 33
55 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
60 <400> PreSequenceString :
MKKALLAAL VMASGSALAV DGGHIDFNGM VQSGTCKGVV VDTGMHSVTT DGVVTLDTAN 60
VTDTFAEVSA TAVGLLPKEF MISVECDPGA PKNAELTMGS ASYANTSGTL NNNMNITVNG 120
IAPAQNVNIA VHNMKNKAGA AEIKQVHMNN SSEVQELTLD AEGKGQYVFN ASYVKAPNSP 180
AVTAGHVTTN ALYTVAYK 198

<212> Type : PRT
<211> Length : 198
65 SequenceName : SEQ ID 34

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKPNMIVGAL ALTSVFMAGH LQAADGTVHF RGEIIDSTCE VTPETKDQVV DLGKVNRTAF 60
SGVDDVAAPT AFSIDLTQCP ETFKSAAIRF DGNEDAHGNG NLAIGTPLDN SNDAAAGISP 120
SDNSGDYTGGA GAVSAAKGVA IRLYNRADNT QVKLYENSAS TPISNGNASM KFMARYIATE 180
10 TTIDPGTANA DSQFTVEYIK 200
<212> Type : PRT
<211> Length : 200
SequenceName : SEQ ID 35
SequenceDescription :

15

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
20 MPIFQREGHL KYSFAAGEYQ AGNYDSASPR FGQLDLIYGL PWGMTAYGGV LISNNYNAFT 60
LGIGKNFGYI GAISIDVTQA KSELNDRDS QGQSYRFLYS KSFESGTDFFR LAGYRYSTSG 120
FYTFQEATDV RSDADSDYNR YHKRSEIQGN LTQQLGAYGS VYLNLTQQDY WNDAGKQNTV 180
SAGYNGRIGK VSYSIAYSWN KSPEWDESDR LWSFNISVPL GRAWSNYRVT TDQDGRTNQQ 240
VGVSGTLLED RNLSYSVQEG YASNGVGNSG NANVGYQGGG GNVNVGYSYG KDYRQLNYSV 300
25 RGGVIVHSEG VTLSQPLGET MTLISVPGAR NARVVNNGGV QVDWMGNAIV PYAMPYRENE 360
ISLRSDSLGD DVDVENAFQK VVPTRGAIVR ARFDTRVGYR VLMTLIRSAG SPVPFGATAT 420
LITDKQNEVS SIVGEEGQLY ISGMPEEGRV LIKWGNDASQ QCVAPYKLSL ELKQGGIIPV 480
SANCQ 485
<212> Type : PRT
30 <211> Length : 485
SequenceName : SEQ ID 36
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSGYTVKPPT GDSNEQTQFI DYFNLFYSKR DQEQISISQQ LGNYGATFFS ASRQSYWNTS 60
RSDQQISFGL NVPFGDITTS LNYSYSNNIW QNDRDHLLAF TLNVPFSHWM RTDSQSAFRN 120
40 SNASYSMSND LKGGMTNLSG VYGTLLPDNN LNYSVQVGNT HGGNTSSGTS GYSTLNYRGA 180
YGNTNVGYSR SGDSSQIYYG MSGGIIAHAD GITFGQPLGD TMVLVKAPGA DNVKIENQTG 240
IHTDWRGYAI LPFATEYREN RVALNANSLA DNVELDETUV TVIPTHGAIA RATEFNAQIGG 300
KVLMTLKYGN KSVPFGAIVT HGENKNGSIV AENGQVYLTG LPQSGKLQVS WGNDKNSNCI 360
VDYKLPEVSP GTLLNQQTAI CR 382
45 <212> Type : PRT
<211> Length : 382
SequenceName : SEQ ID 37
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSALYERSQL TQVMISSAPA TAETMEKAEY LRLDCTIKEV QFTAGQKQDI DVTTLNSTEQ 60
55 ENINGLGASS EISMSGNFYL NQAQNALRDA YDNDTVYAFK VQFPGSGGFK FLAEVRQHTW 120
SSGTNGVVAA TFSLRLKGKP VSYVVPLAFV KNLDKTLTVN TGALLTMSVS VNGGTPPYKH 180
AWKKDGPVE GQTTDTFSKP GAQSGDKGAY TCEVTDSAEQ PQSITSDACT VTVNGAGG 238
<212> Type : PRT
60 <211> Length : 238
SequenceName : SEQ ID 38
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

MRNKPFFYLLC AFLWLAVSHA LAADSTITIR GYVRDNGCSV AAESTNFTVD LMENAAKQFN 60
NIGATTPVVP FRILLSPCGN AVSAVKVGFT GVADSHNANL LALENTVSAA SGLGIQLLNE 120
QQNQIPLNAP SSAISWTTLT PGKPNTLNFY ARLMATQVPV TAGHINATAT FTLEYQ 176

5 <212> Type : PRT
<211> Length : 176
SequenceName : SEQ ID 39
SequenceDescription :

10 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
15 MNKSVVSISA AMLVLLCQPV MGSEISPATP SDEDNYTFDP QLFRGSRFSQ SSLAKLITRE 60
SVAPGNYKMD IYTNNKLSGS WNVTFKEAAD GRVLPCLTPE VADAIGLKTG EDKGEKDPVC 120
TFAKELAPGI TSQTQLSCLR LDLSVPQSQL ISRPRGYVPP SELDTGASLA FMNYIANYYN 180
VAYSGQNAHS QRSLWASFNG GINLGAWQYR QLSNMTWDND KGNQWNNIRS YLQRPLPAIN 240
SQLMMGQLIT SGRFFSGLSY HGVSLATDER MLPDSMRGYA PTIRGVAATN ARVSVMQNGH 300
EIIYQTTVAPG PFEINDLYPT SYSGDLDTV TVTEANGAVSRF SVPFSAVPES MRPGTSRYNV 360
20 EVGKTQDSGD DSMFGDLTWQ HGMTNTLTFTN SGSRIADGYQ ALMLGGVYGS SLGAFCANLT 420
WSHARVPESE AQSGWMSQLT WSKTFQPTST TVSLAGYRYS TSGYRDLADV LGERHAASNK 480
QSWDSSQWRQ QSRFDLTLSQ SLANYGNLFV SGSTQNYRGG KSRDTQLQLG YSNSFSHGSI 540
MNLSVGRQRM GGYKDNSDDM QTVTSLSFSE PLGGNGPRVP SLSNSWTHST DGSSQLQSSL 600
TGMLDEAQT NYSLNVMRDQ QYKQTTLSGN MQKRFSQTTV GLNASKGQDY WQASGNVQGA 660
25 MAVHGGGITF GPYLGETFAL VEAKGAEGAK VYNSSQLEIN DSGYALVPAV TPYRYNRISL 720
DPQGMGDGAE LVDSEKQVAP VAGAAVKVIF RTRPGKALLI KSRMADGSEL PMGADVLDEN 780
NTVVGIAQGQ GQIYLRTEQT KGHLSSVRWGE GANDSCQLPF DISGKDSNSP IIRLNETCQS 840

30 <212> Type : PRT
<211> Length : 840
SequenceName : SEQ ID 40
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MKLAACFLTL LPGFAVAASW TSPGFPAFSE QGTGTFVSHA QLPKGTRPLT LNFDQQCWQP 60
ADAIKLNQML SLQPCSNTPP QWRLFRDGKY TLQIDTRSGT PTLMISIQNA AEPVANLVRE 120
CPKWDGLPLT LDVSATFPEG AAVRDYYSQQ IAIKNGQIT LQPAATSNGL LLLERAETDA 180
SAPFDWNNAT VYFVLTRFE NGDPSNDQSY GRHKDGM AEI GTFHGGDLRG LTNKLDYLQQ 240
LGVNALWISA PFEQIHGWVG GGTGKDFPHY AYHGYTQDW TNLDANMGNE ADLRTLVD SA 300
HQRGIRILFD VVMNHTGYAT LADMQEYQFG ALYLSGDEVK KTLGERWSDW KPAAGQWHS 360
FNDYINFSDK TGWDKWWGKN WIRTDIGDYD NPGFDDLTMS LAFLPDIKTE STTASGLPVF 420
45 YKNKTDTHAK AIDGFTPRDY LTHWLSQWVR DYGDGFRVD TAKHVELPAW QQLKTEASAA 480
LREWKKANPD KALDDKPFWM TGEAWGHGVM QSDYYRHGFD AMINFQYQEQ AAKAVDCIAQ 540
MDTTWQQMAE KLQGFNVLSY LSSHDTLRLR EGGDKAAELL LLAPGAVQIF YGDESSRPFG 600
PTGSDPLQGT RSDMNWQDVS GKSAANVAHW QKISQFRARH PAIGAGKQTT LSLKQGYGFV 660
REHGDDKVLV IWAGQQ 676

50 <212> Type : PRT
<211> Length : 676
SequenceName : SEQ ID 41
SequenceDescription :

55 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
60 MPQRHHQGHK RTPKQLALII KRCLPMVLTG SGMLCTTANA EEEYFDPIML ETTKSGMQTT 60
DLSRFSKKYA QLPQTYQVDI WLNNKKVSQK KITFTANAEQ LLQPQFTVEQ LRELGIKVDE 120
IPALAEKDDD SVINSLEQII PGTAAEFDFN HQRLNLSIPQ IALYRDARGY VSPSRWDDGI 180
PTLFTNYSFT GSDNRYRQGN RSQRQYLNMQ NGANFGPWRL RNYSTWTRND QASSWNTISS 240
YLQORDIKALK SQLLLGESAT SGSIFSSYNF TGVQLASDDN MLPNSQRGFA PTVRGIANSS 300
AIVTIRQNGY VIYQSNVPAG AFEINDLYPS SNSGDLEVTI EESDGTQRRF IQPYSSLPMM 360
65 QRPGLHLYSA TAGRYRADAN SDSKEPEFAE ATAIYGLNNT FTLYGGLLGS EDYYALGIGI 420
GGTLGALGAL SMDINRADTQ FDNQHSFHGY QWRTQYIKDI PETNTNIAVS YYRYTNDGYF 480
SFDEANTRNW DYNRSQKSEI QFNISQTIFD GVSLYASGSQ QDYWGNNEKN RNISVGVSGQ 540

QWGIGYSLNY QYSRYTDQNN DRALSLNLSI PLERWLPRSR VSYQMTSQKD RPTQHEMRID 600
GSLDDGRLS YSLEQSLDDD NNHNSSVNAS YRSPYGTFSY GYSYGNDSSQ YNYGVTGGVV 660
IHPHGVTLISQ YLGNAFALID ANGASGVRIQ NYPGIATDPF GYAVVPYLTT YQENRLSVDI 720
TQLPDNVDLE QTTQFVVPNR GAMVAARFNA NIGYRVLVTV SDRNGKPLPF GALASNDDTG 780
5 QQSIVDEGGI LYLSGISSKS QSWTVRWGNQ ADQQCQFAFS TPDSEPTTSV LQGTACH 838

<212> Type : PRT
<211> Length : 838
SequenceName : SEQ ID 42
10 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
15 <400> PreSequenceString :
MMFRNRILLI FILWANFTWA GCRTTASLNI TDGINVGEIL ANETSFSKSV VFTGISCDTS 60
TDKIVYKNIQ SDWVEVGPFQ NGEKLKVKIE SLGKTSDTIG KSSNAQAVLP YVVKIARGTP 120
DFTGERKSTW FISDTVIANI GGESSSIDF WLKICKALKE NWCNVNLTSL LAGDTFTLGL 180
NISYYPKNTT CKPENTVIKV DDIALFQLRN QGKIAANSKE GTITLKCDNL FGDKKQASRN 240
20 MVVYLSSSDL VKGSNTILRG KTDNGVGFVL DLTEPPKGTE AAIKISANGD QGAATSLWKT 300
DKPGVSLNSN IINIPVMASY YVYDEKKVKS GALEATALIN VKYD 344
<212> Type : PRT
<211> Length : 344
SequenceName : SEQ ID 43
25 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
30 <400> PreSequenceString :
MIKKASLLTA CSVTAFAAWA QDTSPDTLVV TANRFEQPRS TVLAPTTVVT RQDIDRWQST 60
SVNDVLRRLP GVDITQNGGS GQLSSIFIRG TNASHVLVLI DGVRNLNLAGG SGSADLSQFP 120
IALVQRVEYI RGPRSAVYGS DAIGGVNII TTRDEPGTEI SAGWGSNSYQ NYDVSTQQQL 180
GDKTRVTLLG DYAHTHGYDV VAYGNTGTQA QPDNDGFLSK TLYGALEHNF TDAWSGFVRG 240
35 YGYDNRTNYD AYYSPGSPLV DTRKLYSQSW DAGLRYNGEL IKSQILTSYS HSKDYNYPH 300
YGRYDSSATL DEMKQYTVQW ANNIIGHGN VGAGVDWQKQ STAPGTAYVK DGYDQRNTGI 360
YLTGLQQVGD FTFEGAARSD DNSQFGRHGT WQTSAGWEFI EGYRFIASYG TSYKAPNLGQ 420
LYGFYGNPNL DPEKSKQWEG AFEGLTAGVN WRISGYRNDV SLDIDYDDHT LKYYNEGKAR 480
IKGVEATANF DTGPLTHTVS YDYVDARNAI TDTPLLRRAK QQVKYQLDWQ LYDFDWGITY 540
40 QYLGTRYDKD YSSYPYQTVK MGGVSLWDLA VAYPVTSHLT VRGKIANLFD KDYETVYGYQ 600
TAGREYTLTG SYTF 614
<212> Type : PRT
<211> Length : 614
SequenceName : SEQ ID 44
45 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
50 <400> PreSequenceString :
MKNKLLFMML TILGAPGIAA AAGYDLANSE YNFAVNELSK SSFNQAAIIG QAGTNNSAQL 60
RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN DASISQAYG NTAMIIQKGS 120
GNKANITQYG TQKTAIVVQR QSQMAIRVTQ R 151
<212> Type : PRT
55 <211> Length : 151
SequenceName : SEQ ID 45
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
60 <400> PreSequenceString :
MNIFAYLLVL VFSMSMSSSA FASVVMGTGR IIFPGDAKEK TIQLRNTSDQ PYIINIHVED 60
ERGSCKNVPF MPTPQTFRME AAAGQALRLI YTGNNLPQDR ESFVWFSSFSQ LPYLNKNDKS 120
65 QNQLILALTN RVKIFYRPSS IVGKSSDAPK NLTYQVKQNR IEVTNPTGYY VTIRAAELLN 180
NGKKVPLANS VMIAPQSTTE WTLPSGISVA PGAQIHLVTV NDYGVNVTSE HAL 233

<212> Type : PRT
<211> Length : 233
SequenceName : SEQ ID 46
SequenceDescription :

5
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
10 MKRLHKRFLI ATFCALLTAT LQAADVITITV NGRVVAKPCT IQTKEANVNL GDLYTRNLQQ 60
PGSASGWHNI TSLTDCPAE TSAVTAIVTG STDNTGYYKN EGTAENIQIE LRDDQDATLK 120
NGDSKTVIVD EITRNAQFPL KARAITVNGN ASQGTIEALI NVIYTWQ 167
<212> Type : PRT
<211> Length : 167
15 SequenceName : SEQ ID 47
SequenceDescription :

Sequence

20 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRAKLLGIVL TPIAIAISSFA STETLSFTPD NINADISLGT LSGKTKERVY LAEEGGRKVS 60
QLDWKFNNAA IIKGAINWDL MPQISIGAAG WTTLGSRGGN MVDQDWMDS NPWTWDESR 120
HPDTQLNYAN EFDLNIKGWL LNEPNYRLGL MAGYQESRYS FTARGGSYIY SSEEGRDDI 180
25 GSFPNGERAI GYKQRFKMPY IGLTGSYRYE DFELGGTFKY SGWVEASDND EHYDPGKRIT 240
YRSKVKDQNY YSVSVNAGYY VTPNAKVIVE GTWNRVTNKK GNTSLYDHND NTSDYSKNGA 300
GIENYNFITT AGLKYTF 317
<212> Type : PRT
<211> Length : 317
30 SequenceName : SEQ ID 48
SequenceDescription :

Sequence

35 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MFFKRGKILS AGRNLKKS LG IVMLLSVGLL LAGCSGSKSS DTGTYSGSVY TVKRGDITLYR 60
ISRTTGTSVK ELARLNGISP PYTIEVGQKL KLGGAKSSSS TRKSTAKSTT KTASVTPSSA 120
VPKSSWPPVG QRCWLWPTTG KVIMPYSTAD GGNKGIDISA PRGTPIYAAG AGKVYVVGNO 180
40 LRGYGNLIMI KHSYDYITAY AHNDTMLVNN GQSVKAGQKI ATMGSTDAAS VRLHFQIRYR 240
ATAIDPLRYL PPQGSKPKC 259
<212> Type : PRT
<211> Length : 259
SequenceName : SEQ ID 49
45 SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MPTPNPLAPV KGAGTTLWVY NGNGDPYANP LSDNDWSRLA KVKDLTPGEL TAESYDDSYL 60
DDEDADWAAT GQGQKSAGDT SFTLAWMPGE QGQALLAWF NEGDTTRAYKI RFPNGTVDVF 120
RGWVSSIGKA VTAKEVITRT VKVTNVGRPS MAEDRSTVTA ATGMTVTPAS TSVVKGQSTT 180
LTVAFAQPEGA TDKSFRAVSA DTKKATVSVS GMTITVKGVA AGKVNIPVVS GNGEFAAAVAE 240
55 INVTAS 246
<212> Type : PRT
<211> Length : 246
SequenceName : SEQ ID 50
SequenceDescription :

60
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
65 MSALYERSQL TQVMISSAPA TAETMDKAEY LRLDCTIKEV QFTAGQKQDI DVTTLCSSTEQ 60
ENINGLGASS EISMSGNFYL NQAQNALRDA YDNDALYAFK VLFPSGKGFK FLAEVRQHTW 120
SSGTNGVVAA TFSRLKGGP VSFVVPLAFV KNLDKTLTVN TGALLTMSVS ANGGTTPPYKY 180

AWKKDGPVD GQTTDTFSKP GAQSADAGKY TCVVTDSEAK AQSPTSVECT VTVSAAAG 238

<212> Type : PRT
<211> Length : 238
5 SequenceName : SEQ ID 51
SequenceDescription :

Sequence

10 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKSTLALVV MGIVASASVQ AAETYNKDGN KLDVYGKVKV MHYMSDNDSK DGDQSYIRFG 60
FKGETQINDQ LTGYGRWEAE FAGNKAESDT AQQKTRLAFA GLKYKDLGSF DYGRNLGALY 120
DVEAWTDMFP EFGDSSAQT DNFMTKRASG LATYRNTDFF GVIDCLNLTL QYQG'NENRD 180
15 VKKQNGDGFQ TSLTYDFGGS DFAISGAYTN SDRTNEQNLQ SRGTGKRAEA WATGLKYDAN 240
NIYLATFYSE TRKMTPTTGG FANKTQNFEE VAQYQDFDGL RPSLGYVLSK GKDIEGIGDE 300
DLVNYIDVGA TYYFNKNMSA FVDYKINQLD SDNKLININD DIVAVGMTYQ F 351

<212> Type : PRT
20 <211> Length : 351
SequenceName : SEQ ID 52
SequenceDescription :

Sequence

25 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRVKHAVVLL MLISPLSWAG TMTFQFRNPN FGGNPNGAF LLNSAQANNS YKDPSYNDDF 60
GIETPSALDN FTQAIQSQIL GLLSNINTG KPGRMVTNDY IVDIANRDGQ LQLNVTDRKT 120
30 GQTSTIQVSG LQNNSTDF 138
<212> Type : PRT
<211> Length : 138
SequenceName : SEQ ID 53
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MKRKVLAMLV PALLVAGAN AAETYNKDGN KLDLYGKVG LHYFSDDASS DGDMSYARIG 60
FKGETQIADQ FTGYGQWEN IGANGPESDK GNTATRLAFA GFGFGQNGTF DYGRNYGVVY 120
DVEAWTDMFP EFGDITYAGA DNFMNGRANS VATYRNNNGF GQVDGLNFAL QYQGNNEKSG 180
LFDQEGSGNG NGRKLAKENG DGSVCPLPMT LTLV 214
<212> Type : PRT
45 <211> Length : 214
SequenceName : SEQ ID 54
SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNTVTLEGGT FNNNGTLNDV VKIEKNSNAV INNTGSLSTL QLHDGTVNNS GIASARVNAQ 60
GDAVFNNLAG GEARKGAILY NSAVVNNAGT WKMGYQDENN NAGTLDIDDK STFNNSGKLI 120
55 LDNSKNAIRF QGSNANATLY NTGEMTLDAAL LGAGAILYDD GASEFINKGV VDAKVTVAVS 180
TAGATESDAF LWNQDGGVIN FDKDNASAVK FTHNNYVALN DGVMNISGNN AVAMEGDKNA 240
QLVNNGVINL GTEGTTDTGL TGMQLDANAT ADAVIENNGT INIFANDSFA FSVLGTEGHI 300
VNNGTVVIAD GVTGSGLIKQ GDSVNVEGVN GNSGNNTTEVH YTDYTLDPMP NTYTTSPFSE 360
TTDSGSSDGS SNNLNGYIVG TNVDGSAGKL KVNNASMNGV GINTGFAAGT ADTTVSFDNV 420
60 VEGINLTDAD AITSTSVVWT AKGSTDASGN VDVIMSKNAY TDVATDASVN DVAKALDAGY 480
TNNELYTSLN VGTTAELNSA LKQVSGSQAT TVFREARVLS NRFSMLADAA PKVGNGLAFN 540
VVAKGDPRAE LGNNTDYDML ALRKTVDLSE SQSMSLEYGI ARLDGDGAQK AGDNGVTGGY 600
SQFFGLKHQM SFDNGMRWNN ALRYDVHNL DSSRSVAYGDV SKTADTDVKQ QYLELRSEGA 660
KTFEPREGLK ITPYAGVKLR HSLEGGYQER NAGDFNLSMN SGSETAVDSI VGLKLDYAGK 720
65 GGWSANATLE GGPNLSSYSKS QRTASLAGAG SQHFNVDGQ KGGGINSLAS VGVKYSSKES 780
SLNLDAYHWK EDGISDKGVM LNFKKTF 807
<212> Type : PRT

<211> Length : 807
SequenceName : SEQ ID 55
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
10 MLNGISNAAS TLGRQLVGIA SRVSSAGGTG FSVAPQAVRL TPVKVHSPFS PGSSNVNART 60
IFNVSSQVTS FTPSRPAPPP PTSGQASGAS RPLPPIAQAL KEHLAAYEKS KGPEALGFKP 120
ARQAPPPPTS GQASGASRPL PPIAQALKEH LAAYEKS KGP EALGFKPARQ APPPPTSGQA 180
SGASRPLPPI AQALKEHLAA YEKS KGP EALGFKPARQ APPPPTSGQA 240
AYEQSKKG 248
<212> Type : PRT
15 <211> Length : 248
SequenceName : SEQ ID 56
SequenceDescription :

Sequence

20 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 MNKKIHSIAL LVNLGIYGVA QAQEPDTPV SHDDTIVVTA AEQNLQAPGV STITADEIRK 60
NPVARDVSEI IRTMPGVNLT GNSTSGQRGN NRQIDIRGMG PENTLILIDG KPVSSRNSVR 120
QGWRGERDTR GDTSWVPPEM IERIEVLRGP AAARYGNAGAA GGVVNIITKK GSGEWHGSWD 180
AYFNAPEHKE EGATKRTNFS LTGPLGDEFS FRLYGNLDKT QADAWDINQG HQSARAGTYA 240
TTLPAGREGV INKTINGVVR WDFAPLQSL LEAGYSRQGN LYAGDTQNTN SDAYTRSKYG 300
DETNRLYRQN YSLTWNGGWD NGVTTSNWVQ YEHTNSRIP EGLAGGTEGK FNEKATQDFV 360
DNDLDDVMLH SEVNLPIDFL VNQTLTLGTE WNQORMKDL SNTQALTGTN TGGAIDGVSA 420
30 TDRSPYSKAE IFSLFAENNM ELTDSTIVTP GLRFDHHSIV GNNWSPALNI SQGLGDDFTL 480
KMGIARAYKA PSLYQTNPNY ILYSKGQGCY ASAGGCYLQG NDDLKAETSI NKEIGLEFKR 540
DGWLAGITWF RNDYRNKIEA GYVAVGQNAV GTDLYQWDNV PKAVVEGLEG SLNVPVSETV 600
MWTNNITYML KSENKTTGDR LSIIPEYTLN STLSWQARED LSMQTTFTWY GKQPKKYN 660
KGQPAVGPE KEISPYSIVG LSATWDVTKN VSLTGGVDNL FDKRLWRAGN AQTGDLAGA 720
35 NYIAGAGAYT YNEPGRTWYM SVNTHF 746
<212> Type : PRT
<211> Length : 746
SequenceName : SEQ ID 57
SequenceDescription :

40 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
45 MGGRFSLRYK KLSYRFVFLT LAGCSSVGNQ SLKNETQESV KTKIVKGKTT KQDVLASFGE 60
PDSRSLIDGE EQWSYTMYSN QSKATSFIPV VGLLAGGADS QTKSLTVSFK GEKVSTYIFN 120
AGTSNVKTGI F 131
<212> Type : PRT
<211> Length : 131
50 SequenceName : SEQ ID 58
SequenceDescription :

Sequence

55 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKIACLSAL AAVLAFTAGT SVAATSTVTG GYAQSDAQGQ MNKMGGFNLK YRYEEDNSPL 60
GVIGSFYTYE KSRTASSGDY NKNQYYGITA GPAYRINDWA SIYGVVGVGY GKFQTTEYPT 120
YKHDTSYDGF SYGAGLQFNP MENVALDFSQ EQSRIRSDV GTWIAGVGYR F 171
60
<212> Type : PRT
<211> Length : 171
SequenceName : SEQ ID 59
SequenceDescription :

65 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
5 MKSIATLVVC AISGIACVNL SAHAAEGEHT ISLGYAHFQF PGLKDFVKDA TAHNRETFSH 60
FVNRNYFSSL GEYTDGRVSG YEGKDKNPQG INIRYRYEIT DDFGVITSFT WTRSLTNSQT 120
FIDVQSADHT RKIKNPAASA RTDIRANYWS LLAGPSWRVN QYMSLYAMAG MGVAKVSADL 180
KIKDNINSSG GFSESNSTKK TSLAWAAGAQ FNLNESVTLD VAYEGSGSGD WRTSGVTAGI 240
GLKF 244
<212> Type : PRT
<211> Length : 244
10 SequenceName : SEQ ID 60
SequenceDescription :

Sequence

15 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILS AAICLTVSGA PAWASEQQAT LSAGYLHVST NAPGSDNLNG INVKYRYEFT 60
DTLGLVTSFS YAGDRNRQIT RYSDTRWHED SVNRNWF SVM AGPSVRVNEW FSAYAMAGVA 120
YSRVSTFSGD YLRVTDNKGK THDVL TGSD GRHSNTSLAW GAGVQFNPTE SVAIDIA YEG 180
20 SGSGDWRTDG FIVGVGYKF 199
<212> Type : PRT
<211> Length : 199
SequenceName : SEQ ID 61
SequenceDescription :

25 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
30 MRKLYAAILS AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT 60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVNRNWF SVM AGPSVRVNEW FSAYAMAGVA 120
YSRVSTFSGD YLRVTDNKGK THDVL TGSD GRHSNTSLAW GAGVQFNPTE SVAIDIA YEG 180
SGSGDWRTDG FIVGVGYKF 199
<212> Type : PRT
35 <211> Length : 199
SequenceName : SEQ ID 62
SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILS AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT 60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVNRNWF SVM AGPSVRVNEW FSAYAMAGVA 120
45 YSRVSTFSGD YLRVTDNKGK THDVL TGSD GRHSNTSLAW GAGVQFNPTE SVAIDIA YEG 180
SGSGDWRTDG FIVGVGYKF 199
<212> Type : PRT
<211> Length : 199
SequenceName : SEQ ID 63
50 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
55 <400> PreSequenceString :
MRKLYAAILS AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT 60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVNRNWF SVM AGPSVRVNEW FSAYAMAGVA 120
YSRVSTFSGD YLRVTDNKGK THDVL TGSD GRHSNTSLAW GAGVQFNPTE SVAIDIA YEG 180
SGSGDWRTDG FIVGVGYKF 199
60 <212> Type : PRT
<211> Length : 199
SequenceName : SEQ ID 64
SequenceDescription :

65 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :
MVMSQKTLFT KSALAVAVAL ISTQAWSAGF QLNEFSSSGL GRAYSSEGAI ADDAGNVSRN 60
PALITMFD RP TFSAGAVYID PDVNISGTSP SGRSLKADNI APTAWVPNMH FVAPINDQFG 120
WGASITSNYG LATEFNNTYA GGSVGGTTDL ETMNLNLGA YRLNNAWSFG LGFNAVYARA 180
5 KIERFAGDLG QLVAGQIMQS PAGKTPQGQA LAATANGIDS NTKIAHLNGN QWGFQWVAGI 240
LYELDKNNRY ALTYSSEVKI DFKGNYSSDL NRVFNNGYGL IPTATGGATQ SGYLTNLNPE 300
MWEVSGYNRV DPQWAIHYSL AYTSSWSQFQQ LKATSTSGDT LFQKHEGFKD AYRIALGTTY 360
YYDDNWTFR T GIAFDDSPVP AQNRSSISIPD QDRFWLSAGT TYAFNKDASV DVGVSVMHGQ 420
SVKINEGPYQ FESEKAWLF GTNFNYAF 448

10 <212> Type : PRT
<211> Length : 448
SequenceName : SEQ ID 65
SequenceDescription :

15 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAFSQAVSGL NAAATNLDVI GNNIANSATY GFKSGTASFA DMFAGSKVGL GVKVAGITQD 60
20 FTDGTTTNTG RGLDVAISON GFFRLVDSNG SVFYSRNGQF KLDENRNLVN MQGLQLTGYP 120
ATGTPPTIQQ GANPTNISIP NTLMAAKTTT TASMQINLNS SDPLPSVNAF DASNADSYNK 180
KGSVTVFDSQ GNAHDMSVYF VKTGDNNWQV YTQDSSDPTG TAEPAMKLVF NANGVLTSNP 240
TENITTGAIN GAEPATFSL S FLNSMQQNTG ANNIVATTQN GYKPGDLVSY QINDDGTVVG 300
NYSNEQTQLL GQIVLANFAN NEGLASEGDN VWSATQSSGV ALLGTAGTGN FGTLTNGALE 360
25 ASNVDLSEL VNMIVAQRNY QSNAQTIKTQ DQILNTLVNL R 401
<212> Type : PRT
<211> Length : 401
SequenceName : SEQ ID 66
SequenceDescription :

30 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 MSKSTFLHIL ISSIILVALI QSSAWANCTN TQIGQTEDGR TALIEFGKIN MTDYFAPAG 60
SLLATTVPVP TNYTSGGATG SSVLWEC DAT DLPNIYFLVA TNGDDR VGGF YDAGGPDGLS 120
DVYATWFAFV GLKQTMAGVT LGRYWKKVPI TSYATQGT KI QIRLQDIPPL HAELYRISTL 180
PDTSATTSWC GNNNTDSSGV GEAKPSGTIY NCVQPNAYIQ LSGTSGILFG HDEPGEDSSV 240
HWDFWGADNG FGYGMR SANR LYNNATCVAR SATPLVLLPT IAEAQLNAGM ESTGNFNV RV 300
40 ECSNSVQSGI SDTQTALGIQ VSEGAYTAAQ KLGILNSNGG VSALVSDNYD AAEMAKGVGI 360
YISNSAHPDT AMTLVGQPGI AKLTPGGNAA GWYPVFEGAT LEGATHPGYS SYSYSFIARL 420
KKLPNQTVSA GKVRATAYIL VKMQ 444
<212> Type : PRT
<211> Length : 444
SequenceName : SEQ ID 67
SequenceDescription :

45 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 MENNRNFPAR QFHSLTFFAG LCIGITPVAQ ALAAEGQTNA DDTLVVEAST PSLYAPQQA 60
DPKFSRPVAD TTRTMTVISE QVIKQDQATN LTDALKNVP G VGAF FAGENG NSTTGDAIYM 120
RGADTSNSIY IDGIRDIGSV SRDTFNTEQV EVIKGPSGTD YGRSAPT GSI NMISKQPRND 180
55 SGIDASASIG SAWFRRTLD VNQVIGDTA VRLNVMG EKT HDAGRDKVKN ERYGVAP SIA 240
FGLGTANRLY LNYLHVTQHN TPDGGIPTIG LPGYSAPSAG TATLNHSGKV DTHNFYGTDS 300
DYDDSTTDTA TMRFEHDIND NTTIRNTTRW SRVKQDYLMT AIMGGASNIT QPTSDVNSWT 360
WSRTANTKDV SNKILTNQTN LTSTFYTASI GHDVSTGVEF TRETQTN YGV NPVTLP AVNI 420
YHPDSSIHPG GLTRNGANAN GQDTDFAIYA FDTLQITRDF ELNGGIRLDN YHTEYDSATA 480
60 CGGSGRGAIT CPAGVAKGSP VTTVD TAKSG NLVNWKAGAL YHLTENGNVY INYAVSQPP 540
GGNNFALAQ S GSGNSANRTD FKPQKANTSE IGTKWQVLDK RLLLTAA LFR TDIENEVEQN 600
DDGTYSQY GK KRVEGYEISV AGNITPAWQV IGGYTQKAT IKNGKDVAQD GSSSLPYTPE 660
HAFTLWSQYQ ATDDISVGAG ARYIGSMHKG SDGAVGTPAF TEGYWVADAK LGYRVNRNLD 720
FQLNVYNLFD TDYVASINKS GYRYHPGEPR TFLLTANMHF 760
65 <212> Type : PRT
<211> Length : 760
SequenceName : SEQ ID 68

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MQMKKLLPIL IGLSLSGFSS LSQAENLMQV YQCARLSNPE LRKSAADRDA AFEKINEARS 60
PLLPQLGLGA DYTYSNGYRD ANGINSNATS ASLQLTQSIF DMSKWRALTL QEKAAGIQDV 120
TYQTDQOTLI LNTATAYFNV LNAIDVLSYT QAQKEAIYRQ LDQTTQRFNV GLVAITDVQN 180
10 ARAQYDTVLA NEVTARNNLD NAVEQLRQIT GNYYPELAAL NVENFKTDKP QPVNALLKEA 240
EKRNLSLLQA RLSQDLAREQ IRQAQDGHL P TLDLTASSGI SDTSYSGSKT RGAAGTQYDD 300
SNMGQNKVGL SFSLPYQGG MVNSQVKQAQ YNFVGASEQL ESAHRSVVQT VRSSFNNINA 360
SISSINAYKQ AVVSAQSSLD AMEAGYSVGT RTIVDVL DAT TTYLNAKQEL ANARYNYLIN 420
QLNIKSALGT LNEQDLLALN NALSKPVSTN PENVAPQTPE QNAIADGYAP DSPAPVVQQT 480
15 SARTTTSNGH NPFRN 495
<212> Type : PRT
<211> Length : 495
SequenceName : SEQ ID 69
SequenceDescription :

20 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 MTKLKLALG VLIATSAGVA HAEGKFSLGA GVGVEHPYK DYDTDVYPVP VINYEGDNFW 60
FRGLGGGYL WNDATDKLSI TAYWSPLYFK AKDSGDHQMR HLDDRKSTMM AGLSYAHFTQ 120
YGYLRTTLAG DTLDNSNGIV WDMAWLYRYT NGGLTVTPGI GVQWNSNQ N EYYYGVSRKE 180
SARSGLRGYN SNDSWSPYLE LSASYNFLGD WSVYGTARYT RLSDEVTDSP IVDKSWTGLI 240
STGITYKF 248
30 <212> Type : PRT
<211> Length : 248
SequenceName : SEQ ID 70
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MKKTLAAGA VLALSSSFTV NAAENDKPQY LSDWWHQSVN VVGSYHTRFG PQIRNDTYLE 60
YEAFAKKDWF DFYGYADAPV FFGGNSDAKG IWNHGSPLFM EIEPRFSIDK LTINTDLSFGP 120
FKEWYFANNY IYDMGRNKDG RQSTWYMG LG TDIDTGLPMS LSMNVYAKYQ WQNYGAANEN 180
EWDGYRFKIK YFVPITDLWG GQLSYIGFTN FDWGS DLGDD SGNAINGIKT RTNNSIASSH 240
ILALNYDHHW YSVVARYWHD GGQWNDDAEL NFGNGNFNVR STGWGGYLVV GYNF 294
45 <212> Type : PRT
<211> Length : 294
SequenceName : SEQ ID 71
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 MLSTQFN RDN QYQAITKPSL LAGCIALALL PSAAFAAPAT EETVIVEGSA TAPDDGENDY 60
SVTSTSAGTK MQMTQRDIPQ SVTIVSQORM EDQQLQTLGE VMENTLGISK SQADSDRALY 120
YSRGFQIDNY MVDGIPTYFE SRWNLGDALS DMALFERVEV VRGATGLMTG TGNPSAAINM 180
VRKHATSREF KGDVSAEYGS WNKERYVADL QSPLTEDGKI RARIVGGYQN NDSWLD RYNS 240
EKTFFSGIVD ADLGDLTLS AGYEYQRIDV NSPTWGGLPR WNTDGSSNSY DRARSTAPDW 300
AYNDKEINKV FMTLKQRFAD TWQATLNATH SEVEFDSKMM YVDAYVNKAD GMLVGPYSNY 360
60 GPGFDYVGGT GWNSGKRKVD ALDLFADGSY ELFG RQHNL M FGGSYSKQNN RYFSSWANIF 420
PDEIGSFYNF NGNFPQTDWS PQSLAQDDTT HMKSLYAATR VTLADPLHLI LGARYTNWRV 480
DTLTYSMEKN HTTPYAGLVF DINDNWSTYA SYTSIFQPQN DRDSSGKYLA PITGN NYELG 540
LKSDWMNSRL TTTLAIFRIE QDNVAQSTGT PIPGSNGETA YKAVDGTVSK GVEFELNGAI 600
TDNWQLTFGA TRYIAEDNEG NAVNPNLPRT TVKMFTSYRL PVMPELTVGG GVNWQNRVYT 660
65 DTVTPYGTFR AEQGSYALVD LFTRYQVTKN FSLQGNVNNL FDKTYDTNVE GSIVYGAPRN 720
FSITGTYQF 729
<212> Type : PRT

<211> Length : 729
SequenceName : SEQ ID 72
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MARFQFKNRK NNGLIFFISF MVMGEAAIAA PLPQWANAPA VTPVAQLSLQ ESILRAFARN 60
10 PGVTQQAQAI GIGEAQIDEA KSAWYPHVGL TGNAGPSRQT DSSGRLDNNV SYGITLTQLV 120
YDFGKTNNDI NLQTAARDSY RFKLMATLTD VAEKTATAYM EVSRYQALCD AAQRNIHSLE 180
NVYNMAALRA NAGLNSSSDE LQAQTRIAGM RSTLEQYQAQ MASAKAQLAV LTGVQPEAIA 240
APPAELAEQP VSLKNIDYQS IPLVLAAENL RQSAQYGVK TKAQYWPTLS IQGGKTRYQT 300
SDRSYWDDQL QLNVNAPLYQ GGAVSAQVQQ AEGQOKISAS QVEQAKLDVL QRASVAYANW 360
15 TGARGREEAG LAQSESAHKT RDVYQNEYKL GKRSINLLLT VEQDVFOAQS AEINANYDGW 420
VAAVNYAAAV NNLIPLAGIK QGLYNDLPDL K 451

<212> Type : PRT
<211> Length : 451
SequenceName : SEQ ID 73
20 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
25 <400> PreSequenceString :
MAKFTSPFSG IKGRALFSLI FAAPMIHATD TATTKDGETI TVTADANTAT EATDGYQPLS 60
TSTATLTDMF MLDIPQVVNT VSDQVLENQN ATTLDALYN VSNVVQTNLT GGTQDAFVRR 120
GFGANRDGSI MTNGLRVLTP RSFNAATERV EVLKGPASTL YGILDPGLI NVVTKRPEKT 180
FHGSVSATSS SFGGGTGQLD ITGPIEGTQL AYRLTGEVQD EDYWRNFGKE RSTFIAPSLT 240
30 WFGDNATVTM LYSHRDYKTP FDRGTIFDLT TKQPVNVDRK IRFDEPFNIT DGQSDLAQLN 300
AEYHLNSQWT ARFDYSYSQD KYSDNQARVT AYDATTGTLT RRVDATQGST QRMHSTRADL 360
QGNVDIAGFY NEILGGVSYE YYDLLRTDMI RCKNAKDFNI YNPVYGNITSK CTTVSASDSD 420
QTIKQESYSA YAQDALYLTN NWIAVAGIRY QYYTQYAGKG RPFNVNTDSR DEQWTPKLGL 480
VYKLTPSVSL FANYSQTFMP QSSIASYIGD LPPSSNAYE VGAKFELFDG ITADIALFDI 540
35 HKRNVLYTES IGDETIAKTA GRVRSRGVEV DLAGALTENI NIIASYGYTD AKVLEDPDYA 600
GKPLPNVPRH TGSFLTYDI HNMPGNNTLT FGGGGHCVSR RSATNGADYY LPGYFVADAF 660
AAYKMKLQYP VTLQLNVKNL FDKTYTSSI ATNNLGNQIG DPREVQFTVK MEF 713

<212> Type : PRT
40 <211> Length : 713
SequenceName : SEQ ID 74
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRTLQGWLLP VFMLPMAVYA QEATVKEVHD APAVRGSIIA NMLQEHDPNF TLYPYDTNYL 60
IYTQTSDLNK EAIASYDWA E NARKDEVKFQ LSLAFPLWRG ILGPNVSLGA SYTQKSWWQL 120
50 SNSEESSPFR ETNYEPQLFL GFATDYRFAG WTLRDVEMGY NHDSNGRSDP TSSWNRLYT 180
RLMAENGWNL VEVKPYVVG NTDDNPDIK YMGYYQLKIG YHLGDAVLSA KGQYNWNTGY 240
GGAELGLSYP ITHVRLYTQ VYSGYGESLI DYNFNQTRVG VGVMLNDLF 289

<212> Type : PRT
<211> Length : 289
55 SequenceName : SEQ ID 75
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAVQKNVIK ILAGTFALML SGCVTVPDAI KGSSPTPQQD LVRVMSAPQL YVGQEARFGG 60
KVVAVQNQQG KTRLEIATVP LDSGARPTLG EPSRGRIYAD VNGFLDPVDF RGQLVTVVGP 120
ITGAVDGKIG NTPYKFMVMQ ATGYKRWHLT QQVIMPPQPI DPWFYGGRGW PYGHGGWGWY 180
65 NPGPARVQTV VTE 193
<212> Type : PRT
<211> Length : 193

SequenceName : SEQ ID 76
SequenceDescription :

5 Sequence

 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
MRKQWLGICI AAGMLAActs DDGQQQTVSV PQPAVCNGPI VEISGADPRF EPLNATANQD 60
YQRDgKSYKI VQDPSRFIQA GLAAIYDAEP GSNLTASGEA FDPTQLTAAH PTLPIPSYAR 120
10 ITNLANGRMI VVRINDRGPY GNDRVISLSR AAADRLNTSN NTKVRIDPII VAQDGSLSGP 180
GMACTTVAKQ TYALPAPPDL SGGAGTSSVS GPQGDILPVS NSTLKSEDPT GAPVTSSGFL 240
GAPTTLAPGV LEGSEPTPAP QPVVTAPSTT PATSPAMVTP QAASQSASGN FMVQVGAVSD 300
QARAQQYQQQ LGQKFGVPGR VTQNGAVWRI QLGPfANKAE ASTLQQRLOT EAQLQSFITT 360
AQ 362

15 <212> Type : PRT
 <211> Length : 362
 SequenceName : SEQ ID 77
 SequenceDescription :

20 Sequence

 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
MIKRVLVVSM VGLSLVGCVN NDTLSGDVYT ASEAKQVQNV SYGTIVNVRP VQIQGGDDSN 60
25 VIGAIGGAVL GGFLGNTVGG GTGRSLATAA GAVAGGVAGQ GVQSAMNKTQ GVELEIRKDD 120
GNTIMVVQKQ GNTRFSPGQR VVLASNGSQV TVSPR 155

 <212> Type : PRT
 <211> Length : 155
 SequenceName : SEQ ID 78
30 SequenceDescription :

 Sequence

 <213> OrganismName : Escherichia coli O157:H7
35 <400> PreSequenceString :
MSKATEQNDK LKRAIIISAV LHVILFAALI WSSFDENIEA SAGGGGGSSI DAVMVD SGAV 60
VEQYKRMQSQ ESSAKRSDEQ RKMKEQQAEE ELREKQAAEQ ERLKQLEKER LAAQEQQKQA 120
EEAAKQAEELK QKQAEAEAAK AAADAKAKAE ADDKAAEEAA KAAADAKKK AEAEAAKAAA 180
EAQKKAEEAAA AALKKKAEEA EAAAAEARKK AAEEKAAADK KAAEKAAAEK AAADKKAAAE 240
40 KAAADKKAAA AKAAAEKAAA AKAAAEADDI FGELSSGKNA PKTGGGAKGN NASPAGSGNT 300
KNNGASGADI NNYAGQIKSA IESKFYDASS YAGKTCTLRI KLAPDGMLLD IKPEGGDPAL 360
CQAALAAAKL AKIPKPPSQA VYEVFKNAPL DFKP 394

 <212> Type : PRT
 <211> Length : 394
45 SequenceName : SEQ ID 79
 SequenceDescription :

 Sequence

50 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
MMKFKKCLLP VAMLASFTLA GCQSNADDHA ADVYQTDQLN TKQETKTVNI ISILPAKVAV 60
DNSQNKRNAQ AFGALIGAVA GGVIGHNVGS GSNSGTTAGA VGGGAVGAAA GSMVNDKTLV 120
EGVSLTYKEG TKVYTSTQVG KECQFTTGLA VVITTTYNET RIQPNTKCPE KS 172

55 <212> Type : PRT
 <211> Length : 172
 SequenceName : SEQ ID 80
 SequenceDescription :

60 Sequence

 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
65 MLLSIITVAF RNLEGIVKTH ASLAHLAQAE DISFEWIVVD GGSNDGTREY LENLNGIYNL 60
RFVSEPDNGI YDAMNKGIAA AQGKFALFLN SGDIFHQDAA YFVRKLKMQK DNVMITGDAL 120
LDFGDGHKIK RSAKPGWYIY HSLPASHQAI FFPVSGLKKW RYDLEYKVSS DYALAAKMYK 180

AGYAFKKLNG LVSEFSMGGV STTNMELCA DAKKVQRQIL HVPGFWAELS WHLRQRTTSK 240
TKALYNKS 248
<212> Type : PRT
<211> Length : 248
5 SequenceName : SEQ ID 81
 SequenceDescription :

Sequence

10 <213> OrganismName : Haemophilus influenzae Rd
 <400> PreSequenceString :
 MKLTTLQTLK KGFTLIELMI VIAIIAILAT IAIPSYQNYT KKAHVSELLQ ASAPYKADVE 60
 LCVYSTNETT SCTGGKNGIA ADIKTAKGYV ASVITQSGGI TVKGNGTLAN MEYILQAKGN 120
 AAAGVTWTTT CKGTDASLFP ANFCGSVTK 149
15 <212> Type : PRT
 <211> Length : 149
 SequenceName : SEQ ID 82
 SequenceDescription :

20 Sequence

 <213> OrganismName : Haemophilus influenzae Rd
 <400> PreSequenceString :
 MLNKKFKLNF IALTVAAYALT PYTEAALVRD DVDYQIFRDF AENKGRFSVG ATNVEVRDKN 60
25 NHSLGNVLPN GIPMIDFSVV DVDKRIATLI NPQYVVGVKH VSNGVSELHF GNLNGNMNNG 120
 NAKSHRDVSS EENRYFSVEK NEYPTKLNGK AVTTEDQTQK RREDYMPRL DKFVTEVAPI 180
 EASTASSDAG TYNDQNKYPA FVRLGSGSQF IYKKGDNYSI ILNNHEVGGN NLKLVGDAYT 240
 YGIAGTPYKV NHENGLIGF GNSKEEHSDP KGILSQDPLT NYAVLGDSGS PLFVYDREKG 300
 KWLFLGSYDF WAGYNKKSQW EWNLYKPEFA KTVLDKDTAG SLTGSNTQYN WNPTGKTSVI 360
30 SNGSESLNVD LFDSSQDTS KKNNHGKSVT LRSGTLTLN NNIDQGAGGL FFEGDYEVKG 420
 TSDSTTWKGA GVSVDGKTV TWKVHNPBSD RLAKIGKGTI IVEGKGENKG SLKVGDTGTI 480
 LKQQADANNK VKAFSQVGIV SGRSTVVLND DKQVDPNSIY FGFRGGRLDA NGNNLTFEHI 540
 RNIDDGARLV NHNTSKTSTV TITGESLITD PNTITPYNID APDEDNPFYF RRIKDGGLY 600
 LNLENYTYYA LRKGASTRSE LPKNSGESNE NWLYMGKTSI EAKRNVNHI NNERMNGFNG 660
35 YFGEEGKNN GNLNVTFKKG SEQNRFLTGT GTNLNGDLKV EKGTLFLSGR PTPHARDIAG 720
 ISSTKKDQHF AENNEVVVED DWINRNFAT NINVTNNATL YSGRNVANIT SNITASDNAK 780
 VHIGYKAGDT VCVRSYDTGY VTCTTDKLSI KALNSFNATN VSGNVNLSGN ANFVLGKANL 840
 FGTISGTGNS QVRLTENSHT HLTGDSNVNQ LNLDKGHIHL NAQNDANKVT TYNTLTVNSL 900
 SGNGSFYYLT DLSNKQGDV VVTKSATGNF TLQVADKTGE PTKNELTLFD ASNATRNNLN 960
40 VSLVGNTVDL GAWKYKLNRV NGRYDLYNPE VEKRNQTVDT TNITTPNNIQ ADVPSVPSNN 1020
 EEIARVETPV PPPAPATPSE TTETVAENSK QESKTVEKNE QDATETTAQN GEVAEEAKPS 1080
 VKANTQTNEV AQSGSETEET QTTEIKETAK VEKEEKAKVE KDEIQEAPQM ASETSPKQAK 1140
 PAPKEVSTDV KVEETQVQAQ PQTQSTTVAA AEATSPNSKP AEETQPSEKT NAEPVTPVVS 1200
 KNQTEINTDQ PTEREKTAKV ETEKTQEPPQ VASQASPKQE QSETVQPQAV LESENVPTVN 1260
45 NAEEVQAQLQ QTTSATVSTK QPAPENSINT GSATAITETA EKSDKPQET AASTEDASQH 1320
 KANTVADNSV ANNESSDPK SRRRRSISQP QETSAEETTA ASTDETTIAD NSKRSPNRR 1380
 SRRSVRSEPT VTNGSDRSTV ALRDLTSTNT NAVISDAMAK AQFVALNVGK AVSQHISQLE 1440
 MNNEGQYNVW VSNTSMNENY SSSQYRRFSS KSTQTQLGWD QTISNNVQLG GVFTYVRNSN 1500
 NFDKASSKNT LAQVNFYSKY YADNHWYLG I DLGYGKFQSN LKTNHNAKFA RHTAQFGLTA 1560
50 GKAFNLGNFG ITPIVGVRYV YLSNANFALA KDRIKVNPIV VKTAFQAQVDL SYTYHLGEFS 1620
 VTPILSARYD TNQSGSKINV NQYDFAYNVE NQQQYNAGLK LKYHNVKLSL IGGLTKAKQA 1680
 EKQKTAELKL SFSF 1694
 <212> Type : PRT
 <211> Length : 1694
55 SequenceName : SEQ ID 83
 SequenceDescription :

Sequence

60 <213> OrganismName : Haemophilus influenzae Rd
 <400> PreSequenceString :
 MALVNKIKTL SSVGILAATL FLAGCQAQSN ILAFTPPAPS ASMNVNRTAV VSVTTKDSRA 60
 IQEIASYTKH GELIKLNASP SVTQLFQQVM QQNLISKGFR VGQLNGSNAW VTVDVREFGT 120
 QVEQGNLRYK LNTKIQATVY VQAKGSYNK SFNVTHSQEG VFNAGNDEIH KVLSQTFNDI 180
65 VNNIYQDQEV AAAINQYSN 199
 <212> Type : PRT
 <211> Length : 199

SequenceName : SEQ ID 84
SequenceDescription :

Sequence

5 -----

<213> OrganismName : Haemophilus influenzae Rd

<400> PreSequenceString :

MLCWIGYKNG ILPQQNSTLY PVLNPSKCGV IFDGFQLVGD DFNSDQTAEN TSPAQQVLYT 60

THLQSCSPIH SGENFAPIPL YKQLKNQPHL SQDLIKWQEN WQACDQLQMN GAVLEQQSLA 120

10 EISDHQSTLS KHGRYLAQEI EKETGIPTY YLYRVGGQSL ESEKSRCCPS CGANWALKDA 180

IFDTFHFCKD TCRLVSNLSW NFL 203

<212> Type : PRT

<211> Length : 203

SequenceName : SEQ ID 85

15 SequenceDescription :

Sequence

<213> OrganismName : Haemophilus influenzae Rd

20 <400> PreSequenceString :

MGAFAFASVT NANIYAEGDI GLSQTANGS NNTRVGPRVS VGYKVGNTRV AGDYTHHGKV 60

DGTKIQGLGA SVLYDFDTNS KVQPYVGARV ATNQFKYTNR AEQKFKSSSD IKLGYGVVAG 120

AKYKLDGNWY ANGGVEYNRL GNFDSTKVN YGAKVGVGYG F 161

<212> Type : PRT

25 <211> Length : 161

SequenceName : SEQ ID 86

SequenceDescription :

Sequence

30 -----

<213> OrganismName : Haemophilus influenzae Rd

<400> PreSequenceString :

MKKLLIASLL FGTTTTVFAA PFVAKDIRVD GVQGDLEQQI RASLPVRAGQ RVTDNVDANI 60

VRSLEFVSGRF DDVKAHQEGD VLVSVVAKS IISDVKIKGN SIIPTEALKQ NLDANGFKVG 120

35 DVLIREKLNE FAKSVKEHYA SVGRYNATVE PIVNTLPNNR AEILIQINED DKAKLASLTF 180

KGNESVSSST LQEQMELQPD SWWKLWGNKF EGAQFEKDLQ SIRDYYLNG YAKAQITKTD 240

VQLNDEKTKV NVTIDVNEGL QYDLRSARII GNLGMSAEL EPLLSALHLN DTFRRSDIAD 300

VENAIKAKLG ERGYGSATVN SVPDFDDANK TLAITLVVDA GRRLTVRQLR FEGNTVSADS 360

TLRQEMRQOE GTWYNSQLVE LGKIRLDRTG FFETVENRID PINGSNDEV VVYKVKERNT 420

40 GSINFGIGYG TEGISYQAS VKQDNFLGTG AAVSIAGTKN DYGTSVNLGY TEPYFTKDG 480

SLGGNVFFEN YDNSKSDTSS NYKRTTYGSN VTLGFPVNEN NSYYVGLGHT YNKISNFALE 540

YNRNLYIQSM KFKGNGIKTN DFDLSFGWNY NSLNRGYFPT KGVKASLGGR VTIPGSDNKY 600

YKLSADVQGF YPLDRDHLWV VSAKASAGYA NGFGNKRLPF YQTYTAGGIG SLRGFAYGSI 660

GPNAIYAEHG NGNGTFKKIS SDVIGGNAIT TASAELIVPT PFVSDKSQNT VRTSLFVDAA 720

45 SVWNTKWKSD KSGLDNMVLK SLPDYGKSSR IRASTGVGFQ WQSPIGPLVF SYAKPIKKYE 780

NDDVEQFQFS IGGSF 795

<212> Type : PRT

<211> Length : 795

SequenceName : SEQ ID 87

50 SequenceDescription :

Sequence

<213> OrganismName : Haemophilus influenzae Rd

55 <400> PreSequenceString :

MLKKTSLIFT ALLMTGCVQN ANVTPQAQK MQVEKVDKAL QKGEADRYLC QDDRVRVHVH 60

ATHKKYKKNL HYVTVTTFQGV SEKLTLMISE RGKNYANIRW MWQERDDFST LKTNLGEILA 120

TQCVSQTSE LSGQ 134

<212> Type : PRT

60 <211> Length : 134

SequenceName : SEQ ID 88

SequenceDescription :

Sequence

65 -----

<213> OrganismName : Haemophilus influenzae Rd

<400> PreSequenceString :

| | | | | | | | |
|----|--|------------|-------------|------------|-------------|-------------|------|
| | MRIIIIFFMG | LNMTNFRLER | ACLFYAWAN | GRCCLSSTN | QPTNQPTNQ | TNQPTNQPTN | 60 |
| | QPTNQNSNVS | EQLEQINVSG | STENSDTKTP | PKIAETVKTA | KTLEREQANN | IKDIVKYETG | 120 |
| | VTVEAGRFG | QSGFAIRGVD | ENRVAINIDG | LRQAETLSSQ | GFKELFEGYG | NFNNTNRNGAE | 180 |
| | IETLKEVNIT | KGADSIKNGS | GSLGGSVIYK | TKDARDYLIN | KDYYVSYKKG | YATENNQSFD | 240 |
| 5 | TLTLAGRYKK | FDVLVVTTSR | NGHELENYGY | KNYNDKIQK | KREKADPYKI | EQDSTLLKLS | 300 |
| | FNPTENHRFT | FAADLYEHRS | RGQDLSYTLK | YQSGNETPE | VDSRHTNDKT | KRNISFSYE | 360 |
| | NFSQTPFWD | LKLTYSQRI | KTRARTDEYC | DAGVRHCEGT | DNPTGLKVTN | GKITRRDGSD | 420 |
| | LQFEEKNNTA | KSSDKTYDFK | KFIDTDKRV | DDKLVLNNPS | DTWYDCSIFN | CENNAKIKVF | 480 |
| | KGNNYYGYDG | KWKEVDLEIK | ELNGKKFAKI | KDNDRKIKSI | LPSSPGYLER | LWQERDLDTN | 540 |
| 10 | TQQLNLDLTK | DFKIWHIEHN | LQYGGSYNTA | MKRMVNRAGN | DASDVQWWAT | PTLGEDSWTG | 600 |
| | KPHTCATTYE | WNANLCPDVD | PEFSYLLPIK | TTGKSVYLF | NFVITDYLSE | DLGYRYDNIH | 660 |
| | YQPKYKHGIT | PKLPDDIVKG | LFIPLPNNNS | SDPNKVKENV | QQNIDYIAKQ | NKKYKAHSYS | 720 |
| | FVSTIDPTSF | LRLQLKYSKG | FRTPTSDEMY | FTFKHPDFTI | LPNTDLKPEI | AKTKEIAFTL | 780 |
| | HNDDWGFIST | SLFKTNYKNF | IDLIFKKQET | FVVGSGRGE | TLPFSLYQNI | NRDNASLKI | 840 |
| 15 | EINSKVFLGK | MAKFMDFNL | SYKYTYQKGR | MNGNIPMNAI | QPRTMVYGLG | YDHPNHKFGF | 900 |
| | DFYTTHVASK | NPEDTYNMFY | KEENKKDSTI | KWRSKSYTIL | DLIGYVQPIK | NLTIRAGVYN | 960 |
| | LTNRKYITWD | SARSIRSFGT | SNVIDQSTGL | GINRFYAPGR | NYKMSVQFEF | | 1010 |
| | <212> Type : PRT | | | | | | |
| | <211> Length : 1010 | | | | | | |
| 20 | SequenceName : SEQ ID 89 | | | | | | |
| | SequenceDescription : | | | | | | |
| | Sequence | | | | | | |
| | ----- | | | | | | |
| 25 | <213> OrganismName : Helicobacter pylori J99 | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | MTYRNGKIDL | KERFSKNRSF | KGIIKKIAKK | YTIKNSLSII | YSLKTHSNSS | LSINKKIFLG | 60 |
| | LGFVSALSAQ | SEDYNSSVYW | LNSVNENNNN | KSYIISPLRT | WAGGNRSFTQ | NYNNSQLYIG | 120 |
| | TKNASATPNH | SSVWFGEKGY | IGFITGVFKA | RDIFITGAVG | SGNELKTGGG | AILVFESSNE | 180 |
| 30 | LTTNGAYFQN | NRAGTQTSWI | NLISNNSVNL | TNTDFGNQTP | NGGFNVMMGRK | ITYNGGSVNG | 240 |
| | GNFGFDNVDS | NGATTISGVT | FNNNGALTYK | GGNGIGGSIT | FTNSNINHYK | LNLNANSVTF | 300 |
| | NNSTLGSMPP | GNANTIGNAY | ILNANNITFN | NLTFNGGWV | FNRSDAHVNF | QGTITINNPT | 360 |
| | SPFVNMTGKV | TINPNAIFNI | QNYTPTIGNA | YTLFSMKNGN | IAYDDVNNLW | NIIRLKNTQA | 420 |
| | TKDNSKNATS | NNNTHTYVY | YNLGGTLYHF | RQIFSPDSIV | LQSVYYGANN | LYYTNSVNIH | 480 |
| 35 | DNVFNLNKN | DDRADTIFYL | NGLNTWNYTQ | ARFAQTYGK | NSALVFNATT | PWANGAIPKS | 540 |
| | NSTVRFGGYE | GVNWGKTGYI | TGTFTADRVY | ITGNMMSGNG | AQTGGGATLN | FVGATEINIA | 600 |
| | GATFKNLKT | SQNSYMTFMA | LNGSGSGSKI | NVSQSDFYDW | TDGGYDFTGN | GVFDSVNFNK | 660 |
| | AYYKFQGAEN | SYNFKNTNFI | AGNFKFQCKT | TIEKSVLNDA | SYAFDGVNNA | FNEKDFNGGS | 720 |
| | FNFNHAEQTN | AFNNNSFSGG | SFSFNKQVD | FNGNSFNGGV | FNFNNTPKAS | FTNDTFNVNN | 780 |
| 40 | QFKINGAQTD | FTFSKGVVFN | MQGLLSSLSV | GTTYQLLNK | SVGYKDNNA | LYQMLRWTSG | 840 |
| | ENPSGKLVDE | NKTAPNSAKI | YNVQFTDNG | TYIYKENFNN | GITLTRLCTL | GYTHCVNIDN | 900 |
| | DAFNLKNVNN | NASNTVFYLN | GMTTWKTAGT | GVFTQDYSGT | NSVLVFNQTT | PFLAGANPTS | 960 |
| | NSVVGFGKTS | GAEWGLVGYI | QGVFKANQID | ITGTIRSGNG | AKTGGGATLV | FNAQERLNIA | 1020 |
| | NANLNNDKAG | LQNSWMNFIV | NNGNLNVNTA | NFSNQTTPHG | FNLKANNITW | DKGSVSGGGN | 1080 |
| 45 | FGVDNANANG | NAVIKNVNF | DNGTLIYKGG | ENSAGNSLTL | ENNTFNNSYN | NAKAQNLIFN | 1140 |
| | NNSFNSSGYS | FNDTKNVTFK | GTNTLINSDF | FSRLKGSVSI | DNNSIFNIER | DLTDKTTYTL | 1200 |
| | LSGDNIKYNN | QALADNVFSK | NLWDLIHYDG | EQGTLLRTDN | NTYFVQFTQS | NGQKFVFEET | 1260 |
| | FNPGSITYKY | FTIHSSPFHT | EADSKDIWNQ | VRKQDFDIPG | KTPVCVGVCY | IAPYKNQDLI | 1320 |
| | GSSAFASLNL | FGATVVGTL | LGSAQEKANN | NGGSIWFGKN | NLLYLHGNFN | ATNIFLTNNF | 1380 |
| 50 | NVGNPNAGGG | ATINFNADET | LSADGLNYTN | FQTVAMGLQT | SASQHSWANF | NSKLSMEIKN | 1440 |
| | SNFRDFTWGG | FRFNSGRITF | ENTTFSGWTN | INGATESGSS | YVNMVANTDL | IFTDSILGGG | 1500 |
| | IRYDLKANNI | IFNNTQMVVD | VSKNVNQSSL | NGNVTFNHSR | LSVKPNAAIN | IGGDQTQTTL | 1560 |
| | ENASSLSFYN | DSVANFNGTT | AFNGVSYLNL | NPNAQVSFNQ | ANFNANVTF | YGIPLFGKTP | 1620 |
| | NFGNSVRLIN | FKGDAKFNQA | TLNLRAKNIH | LNFQCASTFE | NNSTMNLAES | SQASFNALSV | 1680 |
| 55 | EGETNFNLNG | SSLLSFNGNS | VFNAPVNFYA | NNSQISFTHS | ATFNADASFD | LGNNSTLNFQ | 1740 |
| | SVLLNSALNL | LGNGGNNLAI | NAKGNFSFGS | QGILNLSYMN | LFGGDKKASV | YDVLQAQNI | 1800 |
| | GLRGNGGYEK | IRFYGIQIEK | ADYSFNNGVH | SWSFTNPLNT | TETITETLHN | NRLKVQISQN | 1860 |
| | GASNNAMFNL | APSLYDYQQN | PYDESENSYN | HTSDKAGTY | LSSSIKGF | NNEIPGTYN | 1920 |
| | QNQPLQALHI | YNQAISKQDL | NMIASLGKEF | LPKVAKLIAS | GALDNLNLNS | PDSFETIFSI | 1980 |
| 60 | LKEYGITLNL | ANWKSLLKII | NNFSNTANYH | FSQGSVLVGA | IKEGQTNNTS | VVWFGGDGYK | 2040 |
| | NPCAVGDNTC | QMFRTNLGQ | LLNSSVPYLG | YINANFKAKN | IYITGTIGSG | NAWGSAGGSAN | 2100 |
| | VSFESATNLV | LNQANIDAQ | TDKIFSYLGK | EGIDKLFGEK | GLGNVLSNIV | YEESLNDNAI | 2160 |
| | PKDLANMIPK | DLGSKTLSSL | LSPTFVNLL | GVSAFKNAIM | EILNSKTVGD | VFGENGLLNA | 2220 |
| | LDPVKRKEID | QMLLEQIQAH | SSGFEEKFIVK | TLGIENVENF | INNWYKQSL | SSFANNFVPG | 2280 |
| 65 | GLNQALDKIG | SSSDAKDLQS | FLDKTTFGDI | LNQMINQAPL | INKLISWLGP | QDLSVLVNI | 2340 |
| | LNSITNPSKE | LLGAISGMGQ | KVLNDLLGEG | VVNKIMSNOV | LGQMINKIIA | DKGFGGVYHQ | 2400 |
| | GLGSILPKSL | QDELKKGGMG | SLLKPKGLHN | LWQKGNFNFV | AKNHVFNNS | LFSNATGGEL | 2460 |

NFVAGKSIIF NGKNTINFQ YQGRLSFVSK DFSNISLDTL NATNGLTLNA SKNDISVQKG 2520
QICVNVLDLCM TAKGKTTQTN SSSSATAPTN ETLEVSANNF AFLGTIKANG LVDFSKVLQN 2580
TTIGTLDLGP NATFKANNLI VNNAFNNNSN YRANISGNFN VAKGATFSTN ENGLNVGGNF 2640
NSEGPLIFNL NNPTHQTIIN VTGTSTIMSY NNQALINFNT QLKQGAYTLI NANRMVYGYD 2700
5 NQTILGGSLS DYLLKLYTLID FNGKRMQLNG DLSYDNQPV SIKDGGLVVS FKDNQGMVY 2760
SSILYDKIQV TVSDKPMISQ APSLEYVVKR IQGSAGLNAI KSAGNNSIMW LSELF AAKGG 2820
NPLFAPYYLQ DNPTEHIVTL MKDITSALGM LSNSNLKNN TDVLQNLTYT QQMSRLAKLS 2880
NFASFSTDF SERLSSLKNQ RFADAVPNAM DVILKYSQRD KLKNNLWATG VGGVSFVENG 2940
TGTLYGVNVG YDRFVRGVIV GGYAAYGYSG FYERITSSKS DNVDVGLYAR AFIKKSELT 3000
10 SVNETWGANK TQISSNDALL SMINQSYKYS TWTTNAKVNY GYDFMFKNKS IILKPQIGLR 3060
YYYIGMSGLE GVMNNVLYNQ FKANADPSKK SVLTIDFALE NRHYFNTNSY FYAIGGVGRD 3120
LLVNSMGDKL VRFIGNNTLS YRKGDLYNTF ANITTTGGEVR LFKSIFYANAG VGARFGLDYK 3180
MIDIIGNIGM RLAF 3194
<212> Type : PRT
15 <211> Length : 3194
SequenceName : SEQ ID 90
SequenceDescription :

Sequence
20 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKQFKKKPKK IKRSHQNQKT ILKRPLWLMP LLIGGFASGV YADGTDILGL SWGEKSQKVC 60
VHRPWYAIWS CDKWEKTQQ FTGNQLITKT WAGGNAANY HSQNNQDITA NLKNDNGTYF 120
25 LSGLYNYTGG EYNGGNLDIE LGSNATFNLG ASSGNSFTSW YPNGHTDVT SAGTINVNNS 180
VEVGNRVGSG AGTHTGTATL NLNANKVTIN SNISAYKTSQ VNVGNANSVI TINSVSLNGD 240
TCSSLARVGV GANCSTSGPS YSFKGTTNAT NTTFSSSSGS FTFEENATFS GAKLNGGFT 300
FNKKFNATNN TAFNSGSFTF KGTSSFNAN FSNASYTFNN QATFQNSSFN GGTFTFNDQT 360
NQSTQHPQIQ NSSFSGSATT LKGFATFEQA FNNSNHQLTI QNASFNATF NNTGKITIEK 420
30 DASFNNTSFN TPVDTNMTI SGGVTLGSKN DLKNGATLDF GSSKITLTQG TTFNLTSLS 480
EKSVTILNSR GGITYNHLLN HAINSLTNAL KTNESSSKPQ SFAQGLWDMI TYNGVTGQLL 540
NENAATSKPT DSSPSKSTN STQVYQVGYK IGDITYKLQE TFSHNSIIQ ALESGTYTPP 600
PVINGSKFDL SASNYINADM PWYNHKYIP KSQNFTESGT YYLPSVQIWG SYTNSFKQTF 660
SASNSNLVIG YNATWTDHNV SSSDTVAFGD TSGSALNGHC GPWPYYQCTG TTNGTYSAYH 720
35 VYITANLRSG NRIGTGGAAN LIFNGVDSIN IANATITQHN AGAYSSMTF STQNMDSQN 780
LNLNSNGKL LVYGTTFTNQ AKDGKFIENA GQATFENTNF NGGSYQFSGD SLNFSNNNQF 840
NSGSFEIGAK NTIFNNANFN NSTSFNFNNS SATTSFVGDF TNANSNLQIA GNAVFGNSTN 900
GSQNTANFNN TGSVNIAGNA TFDNVVFNSP TNSVKGKVT LNNITLKNLN APLSFGDGTI 960
VPSAHSVINI GEAITNGNPI TLVSSSKAIE YNDAFSKNLW QLINYQGHGA SSEKLVSAG 1020
40 NGVYDVVYSE NNQTYNFQEV FSPNSISIRR LGVGMVFDYV DMEKSDRLYY QNALGFMTYM 1080
PNSYNNNLGN LNNTIYYYDN SIDFYASGKT LFTKAESQT FTGQNSAIVF GAKNIWTSVS 1140
DAPQSNVIIR FGDNKGAGSN DASGHCWNLO CIGFITGHE AQKIYITGSI ESGNRISGG 1200
GASLNFNGLQ GILLTNATLY NRAAGTQSS MNFVSNSANI QAQNSYFIDD TAQNKGNPNF 1260
SFNALNLDFS NSSFRGYVGQ TQSVFKFNAV NAISFTNSSN LSSGLYQMQA KSVLFDNSNL 1320
45 SVSVGTSSIK ANAINLSQNA SINASNHSTL ELQGDNLND TSSLNLQSA INVSNNATIN 1380
DYASLIASNG SHLNFGAVN FNSANITSL SSSSIVFKGA VSLRGQFNLS NNSSLDFOGS 1440
SAITSNTAFN FYDNAFSQSP ITFHQALDIK VPLSLGGLL NPNSSVLNL KNSQLVFSQ 1500
GSLNIANIDL LSDLNGKNR VYNIIQADMN GNWYERINFF GMRINDGIYD AKNQYTSFTN 1560
PLNNALKITE SFPKNNQLSVT LSQIPGIKNT LYNIGSEIFN YQKVYNNANG VYSYSDDAQ 1620
50 VFYLTSSVKG YNPNQSYQA SGSNNTTKNN NLTSSESVIS QTYNAQGNPI SALHVYNKGY 1680
NFSNIKALCQ MALKLYPEIK KILGNDFSL SLSNLKGDAL NQLTKLITPS DWKNINELID 1740
NANNSVVQNF NNGTLIIGAT KIGQTDNTSA VVFGGLGYQK PCDYTDIVCQ KFRGTYLGQL 1800
LESISADLGY IDTTFNAKEI YLTGTLGSGN AWGTGGSASV TFNSQTSLL NQANIVSSQT 1860
DGIFSMGLQE GINKVFNQAG LANILGEVAM QSINKAGGLG NLIVNTLGSD SVIGGYLTPE 1920
55 QKNQTLSQLL GQNNFDNLMN DSGLNTAIKD LIRQKLGFWT GLVGGLAGLG GIDLQNPPEL 1980
IGSMSINDLL SKKGLFNQIT GFISANDIGQ VISVMLQDIV KPSDALKNDV AALGKQMIGE 2040
FLGQDTLNSL ESLLQNNQIK SVLDKVLAAK GLGSIYEQGL GDLPNLGKK GIFAPYGLSQ 2100
VWQKGDFFSN AQGNVFNQNS TFSNANGGTL SFNAGNSLIF AGNNHIAFTN HSGTLNLLSN 2160
QVSNINVTML NASNGLKINA TNNNVSVSQG NLFINASCVQ QSDPTTASAT NPCTTAQNNA 2220
60 SSSNASNNAP IALNNNDESL VVTANGFNFS GNIYANGVVD FSKIKGSANV KNLYLYNNAQ 2280
FQANNLTISN QAVLEKNASF VTNNLNIQGA FNNNATQKIE VLQNLVIASN ASLSTGIYGL 2340
EVGGALNNLG AIHFNLNSQ TPVNPLIQVG GIINLNTTQT PFMNVSVANG GTYTLLKSSR 2400
YIDYNINPNS LQSYLKLYTL ININGNHIEE KNGVLTYLQ RVLLQDKGLL LSVALPNSNN 2460
ASQNNILSLS VLHNQIKMSY GNKVMDFTPP TLQDYIVGIQ GQSALNQIEA VGGNNAIKWL 2520
65 STLMMETKEN PLFAPYIYEN HSLNEILGVT KDLQNTASLI SNPNFRNNAT SLEMASYTQ 2580
QTSRLTKLSD FRAREGESNF SERLLELKNK RFSDPNPSEV FVKYSQLSKH PNNLWIQGVG 2640
GASFISGGNG TLYGLNVGYD RLVKSVILGG YVAYGYSGFN GNIMHSLANN VDVGMYARAF 2700

LKRNEFTLSA NETYGGNASH INSSNSLLSV LNQRYNYNTW TTSVNGNYGY DFMFKQKSVV 2760
LKPQVGLSYH FIGLSGMKGK MQNPAYQQFV MHSNPSNESV LTLNMGLESR KYFGKNSYYF 2820
VTARLGRDLL IKAKGDNVVR FVGENTLLYR KGEIFNTFAS VITGGEMHLW RLMYVNAGVG 2880
LKMGLQYQDL NITGNVGMV AF 2902

5 <212> Type : PRT
<211> Length : 2902
SequenceName : SEQ ID 91
SequenceDescription :

10 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :

15 MAFKKARLIS RFISKGSFKL NKISKKFFTL NQILKREKPL KRHKKTKSIE KPFNKKNKSFL 60
KASVLLIGAL GGLSHLRANE CRYWSWSSWS YQDNIESGPN SPTHNSYCLF SSAQGSGETYY 120
LNTLTYSAG GASFTQKFNG GTLDIGGNIR FGGTGINGGD VGYITGTINA QTMNFNSSHI 180
TTGNSYADGG GTTLNFNATN NITINQASFD NSDAGTQKSY MNFKGSNIKI SGSSFTDDTN 240
GGFNFSGNNN NSTISFNQTS FNQGTYNFSN SATLSFNNSN FNQGTYHFNS AQSTFENSNF 300
NQGTYNFNDN TSFNNDTFNQ GTYNFNSSKV SFSGANTLNS SSPFASLKGS VSFNSGAIFN 360
20 LNQTLNNTQT YDILTNGAI QYGVYQSYLW DLINYKGDKA ISHVEVSNT YDVTFDINGQ 420
DETLOETFEN OSIITQFLGD DLQQAQQTQ QEDVANSQNA LNKVASDNTI ANNDTSYTQS 480
SNPTILKDAQ GLENTNQIQ QDEKALEKDL AQIKQLANST TGFNEQAFTQ AQKQEQQDEQ 540
ALQNDENAFN TEQEGLEQAI ANAKHANPTP NPTSPPTPTP IKHTAPNTPP SQVPPTPPSQ 600
NLPKTNVWNG VYWLQNKYYS NKGIYYIDPN LSGQSGQSGN TLSTYTANLL GRSFGVNANN 660
25 GTLIIGNNTE SVNDNGLIWI GHGGFGYITG TFSAAINIYT NNFKTGEGVS NSDGGGANIT 720
FKASDNITMD GLNYNNAETV TKMIQTGASQ HSYTTFDATN NISVTDSDFS DMTWGKFSFS 780
AKNISFSNAS FSGFTNPGGS STISTNASNS LSFTDSRLNG GAIYNLQANS LIFNNTQAVF 840
NVLYSRGTSN FNATTQLLGN TSFTLSSQSL LNFNGDTTLQ NNANITLGNK SQAFAKNSLT 900
LDNNSNLSLD NQSVLNANGT SAFNNQASLN IYNGSQAAFS SLFFNGGTLN LNANSKLNAS 960
30 SASFSNNTTI NLDDSVLNAN NTSSLNANIN FQGASQADFG GNTTIDTASF NFDSASSLNF 1020
NNLTANGALN FNGYAPSLTK ALMNVSQGFV LGNNGDINLS DINIFDNITK SVTYNINLNAQ 1080
KGITGISGAN GYEKILFYGM KIQNATYSN NNIQTWSFIN PLNSSQIIQE SIKNGDLTIE 1140
VLNNPNSASN TIFNIAPELY NYQDSKQNPY GYSYDYSNQ AGTYYLTSNI KGLFTPKGSQ 1200
TPQTPGTYSF FNQPLNSLNI YNKGFSSENK KTLGLGILSN SATLKEMIES NQLDNITNIN 1260
35 EVLQLLDKIK ITQAQKQALL ETINHLTDNI NQTFNNGNLV IGATQDNVTN STSSIWFNGN 1320
GYSSPCALDS ATCSSFRNTY LGQLLGSTSP YLGYINADFK AKSIYITGTI GSSNAFESGG 1380
SADVTQFSAN NLVLNKANIE AQATDNIFNL LGQEGIDKIF NQGNLANVLS QMAMEKIKQA 1440
GGLGNFIENA LSPLSKELPA SLQDETLLGQ IGQNNLDDL NNSGVMNEIQ NIISQKLSIF 1500
GNFVTPSIE NYLAKQSLKS MLDDKGLLNF IGGYIDASEL SSILGVILKD ITNPPTSLQK 1560
40 DIGVVANDLL NEFLGQDVVK KLESQGLVSN IINNVISQGG LSGVYNQGLG SVLPPSLQNA 1620
LKENDLGTL SPRGLHDFWQ KGYFNFLSNG YVFNNSFSF NATGGSILNFV ANKSIIFNGD 1680
NTIDFSKYQG ALIFASNGVS NINITTLNAT NGLSLNAGLN NVSVQKGEIC INLANCPTTK 1740
NSSPANSSVT PTNESLSVHA NNFTFLGTII SNGAIDLSQV TNNSVIGTLN LNENATLQAN 1800
NLITITNAFNN ASNSTANIDG NFTLNQQATL STNASGLNVM GNFNSYGLDV FNLSHVSVA 1860
45 IINTQGTATI MANNNPLIQF NASSKEVGTY TLIDSAKAIY YGYNNQITGG SSLDNYLKLY 1920
ALIDINGKHM VMTDNGLTYN GQAVSVKDGG LVVGFKDSQN QYIYTSILYN KVKIAVSNDP 1980
INNPQAPTLK QYIAQIQGVQ SVDSIDQAGG NQAINWLNKI FETKGSPLFA PYLESHSTK 2040
DLTTIAGDIA NTLEVIANPN FKNDATNLIQ INTYTQQMSR LAKLSDTSTF ARSDFLERLE 2100
ALKNKRFAA IPNAMDVILK YSQNRVKNV VWATGVGGAS FISGGTGTLY GINVGYDRFI 2160
50 KGVIVGGYAA YGYSGFHANI TQSGSSNVN GVYSRAFIK SELTMSLNET WGYNKTFINS 2220
YDPLLSIINQ SYRYDTWTTD AKINYGYDFM FKDKSVIFKP QVGLSYYYIG LSGLRGIMDD 2280
PIYNQFRANA DPNKKSULTI NFALESRYHF NKNSYFVIA DVGRDLFINS MGDKMVRFIG 2340
NNTLSYRDGG RYNTFASIT GGEIRLFKTF YVNAGIGARF GLDYKDINIT GNIGMRYAF 2399

55 <212> Type : PRT
<211> Length : 2399
SequenceName : SEQ ID 92
SequenceDescription :

60 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :

65 MEIQQTHRKI NRPLVSLVLA GALISAIPQE SHAAFFTTVI IPAIVGGIAT GTAVGTVSGL 60
LSWGLKQAE ANKTPDKPKD VWRIQAGKGF NEFPNKEYDL YKSLSSKID GGWDWGNAAR 120
HYWVKGGQWN KLEVDMKDAV GTYKLSGLRN FTGGDLVDNM QKATLRLGQF NGNSFTSYKD 180
SADRTTRVNF NAKNISIDNF VEINNVRVSG AGRKASSTVL TLQASEGITS SKNAEISLYD 240

GATLNLASNS VKLNGNVWVG RLQYVGAYLA PSYSTINTSK VQGEVDFNHL TVGDQNAQA 300
GIIASNKTHI GTLDLWQSAG LNIIAPPEGG YKDKPNSTTS QSGTKNDKKE ISQNNNSNTE 360
VINPPNNTQK TETEPTQVID GPFAGGKDTV VNIFHLNTKA DGTIKVGGFK ASLTTNAAHL 420
NIGKGGVNL NQASGRLLV ENLTGNITVD GPLRVNNQVG GYALAGSSAN FEFKAGVDTK 480
5 NGTATFNNDI SLGRFVNLKV DAHTANFKGI DTGNGGFNTL DFSGVTDKVN INKLITASTN 540
VAVKNFNINE LIVKTNGISV GEYTHFSEDI GSQSRINTVR LETGTRSIFS GGVKFKSGEK 600
LVINDFYSP WNYFDARNVK NVEITRKFS STPENPWGTS KLMFNNLTG QNAVMDYSQF 660
SNLTIQGDFI NNQGTINYLV RGGKVATLNV GNAAAMMFNN DIDSATGFYK PLIKINSAQD 720
LIKNTHEVLL KAKIIGYGNV STGTNGISNV NLEEQFKERL ALYNNNNRMD TCVVRNTDDI 780
10 KACGMAIGNQ SMVNNPDNYK YLIGKAWRNI GISKTANGSK ISVYVLGNST PTENGGNTTN 840
LPTNTTNNAH SANYALVKNA PFAHSATPNL VAINQHDFGT IESVFELANR SKDIDTLYTH 900
SGAQGRDLLQ TLLIDSHDAG YARQMIDNTS TGEITKQLNA ATDALNNVAS LEHKQSGLOT 960
LSLSNAMILN SRLVNLRSKH TNHINSFAQR LQALKGQEFA SLESAAEVLY QFAPKYEKPT 1020
NVWANAIGGA SLNSGSNASL YGTSAGVDAF LNGNVEAIVG GFGSYGYSSF SNQANSLSNG 1080
15 ANNANFGVYS RFFANQHEFD FEAQAGALSD QSSLNFKSTL LQDLNQSYNY LAYSATARAS 1140
YGYDFAFFRN ALVLKPSVGV SYNHLGSTNF KSNSQSQUAL KNGASSQHLF NANANVEARY 1200
YYGDTSYFYL HAGVLQEFH FGSNDVASLN TFKINAARSP LSTYARAMMG GELQLAKEVF 1260
LNLGVVYLHN LISNASHFAS NLGMRYSF 1288
<212> Type : PRT
20 <211> Length : 1288
SequenceName : SEQ ID 93
SequenceDescription :

Sequence
25 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKHILSLTL GSLLVSTLSA EDDGFYTSVG YQIGEEAQM V TNTKGIQDLS DRYESLNNLL 60
NRYSTLNTLI KLSADPSAIN AVRENLGASA KNLIGDKANS PAYQAVLLAI NAAVGFWNV 120
30 GYVTQCGGNA NGQKSISST IFNNEPGYRS TSITCSLNGH SPGYYPMSI ENFKKLNEAY 180
QILQTALKRG LPALKENNGK VNVTYTYTCS GDGNMNCSSQ VTGVNNQKDG TKTKIQTIDG 240
KSVTTTISST VVDSRADGNT TGVSYTEITN KLEGVPDSAQ ALLAQASTLI NTINNACPYF 300
HASNSSEANA PKFSTTTGKI CGAFSEEISA IQKMITDAQE LVNQTSVINE HEQTTPVGNN 360
NGKPFNPFTD ASFAQGM LAN ASAQAKMLNL AEQVGQAINP ERLSGTFQNF VKGFLATC 420
35 PSTAGTGGTQ GSAPGTVTTQ TFASGCAYVG QTITNLKNSI AHFGTQEQQI QQAENIADTL 480
VNFKSRYSSEL GNTYNSITTA LSNIPNAQSL QNAVSKKNNP YSPQIDTNY YLNQNSYNQI 540
QTINQELGRN PFRKVGIVSS QTNNGAMNGI GIQVGYKQFF GQKRKGARY YGFFDYNHAF 600
IKSSFFNSAS DVWTYGFAD ALYNFINDKA TNFLGKNNKL SVGLFGGIAL AGTSWLNSEY 660
VNLATMNNVY NAKMNVANFQ FLENMGVRMN LARPKKSD HAAQHGIELG LKIPTINTNY 720
40 YSFMGAELKY RRLYSVYLY VFAY 744
<212> Type : PRT
<211> Length : 744
SequenceName : SEQ ID 94
SequenceDescription :

Sequence
45 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
50 MIKKAKKFIP FFLIGSLLAE DNGWYMSVGY QIGGTQQFIN NKQLLENQNI INSITQSAIN 60
IAGPTTGLIT LSSQTVIDAL GYGVSNTVGN QLEGISNILN QIGKRKDFYS SRQISSISQQ 120
IIGLKGSSDP LKAHSSQITA KLLSNTQSAF DQGIALSSNI ISAVNSLNPS NNSQEVKAQL 180
QNTAQSMAL LQQIEHSITK TTSTTYAQL LSNLTDVNA SSNNTTYVSA LVNALNTLGV 240
GVFPTTSTH VVLNPPGQVV FYPTNSLLGS TSSNSNNQQQ YNNTLLMNTL QGELSTNNQN 300
55 NPNGCANQIQ CLEQFIQNL PLAATPTSTN QANQQVQAIA QKLQSVAINA LDNNAINNTT 360
YNLNNLHNL NFQAYQSTIE QYNNALKQIS WISFSEPKNL LKNTSNYQI GTVTNDQGN 420
ISAYDCTSAT GSLSSDASSG ISCSATSTN NTNSFDNSLV ATSKVQTING KEQIGVNSFN 480
LVSQVWSVYN SLKTSEENLQ KNAKILCNNG SQSGTSPCNS SSGGLSISGN AQLQNILSPT 540
NGTTTNTQAK SNASKLKAMV MVNNEEEAKT TNFNQSSGPT TQSSNSTVMG ALNTVLQNV 600
60 NFQQSIQSAF QNQNENIQAW ANALYNTSNP NGNQSQNLTT NNNQDLRIQL RANFYQLINT 660
INQQVPTDMN ALINQSQQTQ QTSASASTN NACASGMGSS GNWCYQQWSD SKAYYSGLQS 720
ALGYQTQATT QNGSSGGSNI TYNVQQITLT SGGLLNQIIT NLKSVNGGSN GGSSNGTSQ 780
INTAYQMLTD ASDGKLGTYN SSNSNSSNS GNNNGYTPCN STNGSNGTSG SNCYEPNKQ 840
NATTATTTT SNLQKVYND QKIANIIASS GNNKGVENGL KQFFEALKSN SSSLSNLCGN 900
65 GSSGSSSTCS GGLINLLGAI PTNGVSDTNN LINLLTEFIK TAGFIQNKDS NVSTSLTSAF 960
QAITSAISQ FQALQNDISP NAILTLQEI TSNTTTIQSF SQTLRQLLGD KTFMVQQL 1020
IDAMINARNQ VQNAQNQANN YGSQPVLSQY AAKSTQHGM SNGLGVGIGY KYFFGKARKL 1080

GLRHYFFFDY GFSEIGLANQ SVKANIFAYG VGTDFLWNLF RRTYNTKALN FGLFAGVQLG 1140
GATWLSSLRQ QIIDNWNAN DIHSTNFQVA LNFGVRTNFA EFKRFAKKFH NQGVISQKSV 1200
EFGIKVPLIN QAYLNSAGAD VSYRRLYTFY INYIMGF 1237
<212> Type : PRT
5 <211> Length : 1237
SequenceName : SEQ ID 95
SequenceDescription :

Sequence

10 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKQNLKPFKM IKENLMTQSQ KVRFLAPLSL ALSLSFNVPVG AEEDGGFMTF GYELGQVVQQ 60
VKNPGKIKAE ELAGLLNSTT TNNTNINIAQ TGGNVAGTLG NLFMNQLGNL IDLYPTLKTN 120
15 NLHQCGSTNS GNGATAAAAT NNSPCFQGNL ALYNEMVDSI KTLNQNSKN IFQGDNNNTS 180
ANLSNQLSEL NTASVYLTYM NSFLNANNQA GGIFQNNNTQ AYENGVTAAQ IAYVLKQASI 240
TMGPGSGDGA AGAFLDAALA QHVFNSANAG NDLSAKEFTS LVQNIWNNSQ NALTLANNAN 300
ISNSTGYQVS YGGNIDQARS TQLLNNTTNT LAKVTALNNE LKANPWLGNF AAGNSSQVNA 360
FNGFITKIGY KQFFGENKNV GLRYYGFFSY NGAGVGNGPT YNQVNLLTYG VGTDVLYNVF 420
20 SRSFGSRSLN AGFFGGIQLA GDTYISTLRN SPQLASRPTA TKFQFLFDVG LRMNFGILKK 480
DLKSHNQHSI EIGVQIPTIY NTYYKAGGAE VKYFRPYSVY WVGYYAF 527
<212> Type : PRT
<211> Length : 527
SequenceName : SEQ ID 96
25 SequenceDescription :

Sequence

30 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTLILLSL LSFGLHAEDD GFYASAGIRI GEAAQMVKNT KGIQQLSENY EKLNNLLNNY 60
NTLNTLVKLS SDPSAVNDAR DNLGSSTRNL LDVKANSPAY QAVLLALNAA VGLWQVTSYA 120
FTACGPGSNE NANGGIQTFN NVPQNTTTI TCNSYEPGH GGPISTKNYA IINKAYQIIQ 180
KALTANGEGI PVLSNTTTLK DFTINGDKRT GGEPNKKLVY PWSHGKAIST SWNATITAPT 240
35 TENINTTNSA QELLKQASII ITTLNSACPN FQNGGSGYWA GISGNGTMCG MFKNEISAIQ 300
GMIANAQEAQ AQAKIVSENT QNQNSLDAGK PFNPYTDASF AESMLKNAQA QAEILNQAEQ 360
VVKNFKEIPT AFVNDLGVV YEVOGGERRG TNPQQTTSNT WGAGCAYVGQ TITNLKNSIA 420
HFGTQEQQIQ QAENIADTLV NFKSRYSELG NTYNSITTAL SNIPNAQSLQ NAVSKKNPNY 480
SPQGIDTNY LQNSYNQIQ TINQELGRNP FRKVGIVSSQ TNGAMNGIG IQVGKQFFG 540
40 QKRKWGARYY GFFDYNHAFI KSSFFNSASD VWTYGFQADA LYNFINDKAT NFLGKNNKLS 600
VGLFGGIALA GTSWLNSEYV NLATMNNVYN AKMNVANFQF LFNMGVRMNL ARPKKKDSH 660
AAQHGIELGL KIPTINTNY SFGAELKYR RLYSVYLYNV FAY 703
<212> Type : PRT
<211> Length : 703
45 SequenceName : SEQ ID 97
SequenceDescription :

Sequence

50 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MIKKNRTLFL SLALCASISY AEDDGGFFTV GYQLGQVMQD VQNPGGAKSD ELARELNADV 60
TNNILNNTG GNVAGALSNA FSQYLYSLG AYPTKLNGND VSANALLSGA VSGTCAAAG 120
TAGGTTLNTQ SACTAAGYYW LPSLTDRILS TIGSQTNYGT NTNFPNMQQQ LTYLNAGNVF 180
55 FNAMNKALEK NGTATANSTS STSGATGSDG QTYSQQAIQY LQGGQNILNN AANLLKQDEL 240
LLEAFNSAVA ANIGNKEFNS AAFGLVQGI IDQSQLVYNE LTKNTISGSA VNNAGINSNQ 300
ANAVQGRASQ LPNALYNVQV TLDKINALNN QVRSMPLYLPQ FRAGNSRATN ILNGFYTKVG 360
YKQFFGKKRN IGLRYYGFFS YNGASVGFRS TQNNVGLYTY GVGTDVLYNI FRSYQNRSV 420
DMGFFSGIQL AGETFQSTLR DDPNVKLHGK INNTHFQFLF DFGMRMNFQK LDGKSNRHNQ 480
60 HTVEFGVVVP TIYNTYYKSA GTTVKYFRPY SVYWSYGYSF 520
<212> Type : PRT
<211> Length : 520
SequenceName : SEQ ID 98
SequenceDescription :

Sequence

65

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
5 MKKKFLSLTL GSSLVSALSA EDNGFFVSAG YQIGESAQMV KNTKGIQDLS DSYERLNNLL 60
TNYSVLNALI RQSADPNAIN NARGNLNASA KNLINDKKNS PAYQAVLLAL NAAAGLWQVM 120
SYAISPCGPG KDTSKNGGVQ TFHNTPSNQW GGTITITCGTT GYEPGPYSIL STENYAKINK 180
AYQIIQKAFG SSGKDIPALS DTNTELKFTI NKNNGNTNTN NNGEEIVTKN NAQVLLEQAS 240
TIITTLNSAC PWINNGGAGG ASSGSLWEGI YLKGDSGACG IFKNEISAIQ DMIKNAAIAV 300
EQSKIVAANA QNQRNLDTGK TFNPYKDANF AQSMFANAKA QAEILNRAQA VVKDFERIPA 360
EFVKDSLGVG HEVQNGHLRG TPGSTVTDNT WGAGCAYVGE TVTNLKDSIA HFGDQAERIH 420
10 NARNLAYTLA NFSSQYQKLG EHYDSITAAI SSLPDAQSLQ NVVSKKTNPV SPQGIQDNYY 480
IDSNIHSQVQ SRSQELGSPN FRRAGLIAAS TTNGAMNGI GFQVGYKQFF GKNKRWGARY 540
YGFVDYNHTY NKSQFFNASS DVWTYGVGSD LLVNFINDKA TKHNKISFGA FGGIALAGTS 600
WLNSQYVNLV NVNNYYKAKI NTANFQFLFN LGLRMNLARK KHRATDNAAQ HGIELGTKIP 660
TINTNYSSL-GTTLQYRRLY SVYLVNFVAY 690
15 <212> Type : PRT
<211> Length : 690
SequenceName : SEQ ID 99
SequenceDescription :
20 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
25 MKIKKSLFAL SFSLMASLSR AEDDGFYMSV GYQIGEAVQK VKNTGALQNL ADRYDNLSNL 60
LNQYNYLNSL VNLA STPSAI TGAIDNLSSS AINLTSATTT SPAYQAVALA LNAAVGMWQV 120
IAFGISCGPG PNLGPEHLEN GGVRSDNTP NYSYNTGSGT TTTTCNGASN VGPNGILSSS 180
EYQVLNTAYQ TIQTALNQNQ GGGMPALNSS KNMVVNINQT FTKNPTTEYT YPDGNGNYYS 240
GGSSIPIQLK ISSVNDANL LQQAATIINV LTTQNPVNG GGGAWGFGGK TGNVMDIFGD 300
SFNAINEMIK NAQAVLEKTQ QLNANENTQI TQPDNFPYPT SKDTQFAQEM LNRANAQAEI 360
30 LSLAQVADN FHSIQGPQQ DLEECTAGSA GVINDNTYGS GCAVFKETLN SLEQHTAYYG 420
NQVNQDRALS QTILNFKEAL STLGNDSKAI NSGISNLPNA KSLQNMTHAT QNPNSPEGLL 480
TYSLDTSKYN QLQTVQELG KNPFRIGVI NYQNNNGAMN GIGVQAGYKQ FFGKKRNWGL 540
RYYGFFDYNH AYIKSNFFNS ASDVWTYGVG MDALYNFIND KNTNFLGKNN KLSVGLFGGF 600
ALAGTSWLNS QQVNLTMNG IYNANVSASN FQFLFDLGLR MNLARPKKKD SDHAAQHGM 660
35 LGVKIPTINT DYYSFMGAEL KYRRLYSVYL NYVFAY 696
<212> Type : PRT
<211> Length : 696
SequenceName : SEQ ID 100
SequenceDescription :
40 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
45 MKIKKSLFAL SFSLMASLSR AEDDGFYMSV GYQIGEAVQK VKNTGALQNL ADRYDNLSNL 60
LNQYNYLNSL VNLA STPSAI TGAIDNLSSS AINLTSATTT SPAYQAVALA LNAAVGMWQV 120
IAFGISCGPG PNLGPEHLEN GGVRSDNTP NYSYNTGSGT TTTTCNGASN VGPNGILSSS 180
EYQVLNTAYQ TIQTALNQNQ GGGMPALNSS KNMVVNINQT FTKNPTTEYT YPDGNGNYYS 240
GGSSIPIQLK ISSVNDANL LQQAATIINV LTTQNPVNG GGGAWGFGGK TGNVMDIFGD 300
50 SFNAINEMIK NAQAVLEKTQ QLNANENTQI TQPDNFPYPT SKDTQFAQEM LNRANAQAEI 360
LSLAQVADN FHSIQGPQQ DLEECTAGSA GVINDNTYGS GCAVFKETLN SLEQHTAYYG 420
NQVNQDRALS QTILNFKEAL STLGNDSKAI NSGISNLPNA KSLQNMTHAT QNPNSPEGLL 480
TYSLDTSKYN QLQTVQELG KNPFRIGVI NYQNNNGAMN GIGVQAGYKQ FFGKKRNWGL 540
RYYGFFDYNH AYIKSNFFNS ASDVWTYGVG MDALYNFIND KNTNFLGKNN KLSVGLFGGF 600
55 ALAGTSWLNS QQVNLTMNG IYNANVSASN FQFLFDLGLR MNLARPKKKD SDHAAQHGM 660
LGVKIPTINT DYYSFMGAEL KYRRLYSVYL NYVFAY 696
<212> Type : PRT
<211> Length : 696
SequenceName : SEQ ID 101
SequenceDescription :
60 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
65 MHKKVLLALT ASLICQESLF AKDKDYTLGK VSTAGKKDRS DYSGQVNLGY SGITAPKSWQ 60
DEEVKKYTG RTVISNKALT QQANQSIEEA LQNVPLQIR NATGVGAMPT IQIRGFGAGG 120

SGHSDATLML VNGIPVYMAP YAHIELDIFP VTFQAIDRID VIKGGGSVQY GPNTYGGIVN 180
IITKPIPNQW ENQAAERITY WAKARNAGFA APPDKTGDP5 FIKSLGNNLL YNTYVRSGGM 240
INKHVGIQAA ANWVRGQGFR DNSPSSISNY WLDGVYDINE SNGIKAYYQY YDFAIAQPGS 300
LSEQDYKINR FANLRPLNOK GGRSQRFQAV YENRFGDLDR VGGTFSFTYY GQLMTRDFQV 360
5 SSSYNSANMV TCFSEAACRA AGLPAGYNLA VPYYATNYNG WAEVENPVRS INNAFEPKVN 420
LIVNTGKVRQ TFIMGLRFMT TTFLQROLYN TNECATKTSO EGAGFLCEGP NVMSGWKPHI 480
KHGVYRNWNN WRNNYTAVYL SDRIEAWDGR FFIVPGLRYA FVQYNNENAS NWMQIPEKDL 540
RKIKHMNNWM PSTNIGFIPV QGDHNVLTYP NYQRSFVPPQ LDVLSYGGAE YFTQHFDTVE 600
AGARYTYKDK FSFNADYFRI WARDFATGQY SVYTSGPMKG NVRPINGYSQ GVELELYYRP 660
10 IRGLQFHAAF NYIDTRVTSH GPLTDLNGDV LKGTSYNKHf PFVSPFQFIF DARYNWRKTT 720
IGISSYFYSR AYSGISNSAA GGYGMYQYYS GGNNYESVLN SGYQCEAWCM TQHEGLLPWY 780
WVWNIQVSQI FWENGRHRVT GSLQINNIFN MKYYFTGIGS SPAGLQAPAG RSVTAYLNYT 840
F 841
<212> Type : PRT
15 <211> Length : 841
SequenceName : SEQ ID 102
SequenceDescription :

Sequence
20 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTLLLSLS ASSLLNAEDN GFFISAGYQI GEAAQMVKNT GELKKLSDTY ENLSNLLTNTF 60
NNLNQAVTNA SSPSEINAAI DNLKANTQGL IGEKTNSPAY QAVYLALNAA VGLWNVIAYN 120
25 VQCGPGNSGQ QSVTFEGQPG HNSSSINCNL TGYNNGVSGP LSIENFKKLN QAYQTIQQAL 180
KQDSGFVPLD SAGKQVTITI TTQTNGANKS ETTTTTTTTN DAQTLLEQAS KMISVLTTNC 240
PWVNHNGQON GGAPWGLDTA GNVQVVFATE FSAVTSMIKN AQEIVTQAQS LNQQNNQONAP 300
QDFNPYTSAD RAFAQNMLNH AQAQAKILEL ADQMKKDLNT IPSQFITNYL AACHNGGGTL 360
PDAGVTNNTW GAGCAYVEET ITALNNSLAH FGTQAEQIKQ SELLARTILD FRGSLSNLNN 420
30 TYNSITTTAS NTPNSPFLKN LISQSTNPNN PGGLQAVYQV NQSAYSQLLS ATQELGHNPf 480
RRVGLISSQT NNGAMNGIGV QVGYKQFFGE KRRWGLRYYG FFDYNHAYIK SSFFNSASDV 540
FTYGVGTDVL YNFINDKTTK NSKISFGVFG GIALAGTSWL NSQYVNLATF NNFYSAKMNV 600
ANFQFLFNLG LRMNLAKNKK KASDHAAQHG VELGVKIPTI NTNYYSLLGt QLQYRRLYSV 660
YLNyVFAY 668
35 <212> Type : PRT
<211> Length : 668
SequenceName : SEQ ID 103
SequenceDescription :

Sequence
40 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MRKLFIPLLL FSALEANEKN GFFIEAGFET GLEGTQTQE KRHTTTKNTY ATYNYLPTDT 60
45 ILKRAANLFT NAEAlSKLKF SSLSPVRVLY MYNGQLTiEN FLpYNLNNVK LSFTDAQGNT 120
IDLGVIETIP KHSKIVLPGE AFDSLKEAFD KIDPYTLFLP KFEATSTSiS DTNTQRVFET 180
LNNIKTNLIM KYSNENPNNF NTCpYNNNGN TKNDCWQNFt PQTAEEFTNL MLNMIAVLDS 240
QSWGDAILNA PFEFTNSSTD CDSDPskCVN PGVNGRVDTK VDQQYILNKQ GIINNFRKKI 300
EIDAVVLKNS GVVGLANGYG NDGEYGTlGV EAYALDPKKL FGNDLKTINL EDLRTILHEF 360
50 SHTKGYGHNG NMTYQRVpVT KDGQVEKDSN GKPKDSDGLP YNVCSLYGGS NQPAFPSNYP 420
NSIYHNCADV PAGFLGVTAa VWQQLINQNA LPINYANLGS QTNYNLNASL NTQDLANSML 480
STIQKTFVTS SVTNHHFSNA QSFRSPILG VNAKIGYQNY FNDFIGLAYY GIIKYNyAKA 540
VNQKVQQLSY GGGIDLLLDf ITTYSNKNSP TGIQTKRNFS SSFGIFGGLR GLYNSYyVLN 600
KVKGSGNLDV ATGLNYRYKH SKYSVGISIP LIQRKASVVS SGGDYTNSFV FNEGASHFKV 660
55 FFNYGWVF 668
<212> Type : PRT
<211> Length : 668
SequenceName : SEQ ID 104
SequenceDescription :

Sequence
60 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
65 MNKTTIKILM GmALLSSlQA AEaELDEKSK KPKFADRNTF YLGvGYQLSA INTSFSTSSI 60
DKSYFMTGNG FGvVLGGKFV AKTQAVEHVG FRYGLFYDQT FSSHKSyIST YGLEFSGLWD 120
AFNSPKMFLG LEFGlGIAGA TyMPGGAMHG IIAQYLgKEN SLFQLLVKVG FRFGFFHNEI 180

TFGLKFPVIP NKKTEIVDGL SATTLWQRLP VAYFNYIYNF 220
<212> Type : PRT
<211> Length : 220
SequenceName : SEQ ID 105
5 SequenceDescription :

Sequence

<213> OrganismName : Helicobacter pylori J99
10 <400> PreSequenceString :
MKKTKKTILL SLTLAASLLH AEDNGVFLSV GYQIGEAVQK VKNADKVQKL SDVYEQLSKL 60
LANDNGTSSK TSAQAINQAV NNLNESAKTL AGGTTNSPAY QATLLALRSA LGLWNSMGYA 120
VVCGGYIKKP GENNQKNFHY TDENGNGTTI NCGGSTNSNG THSPNGTNTL KADKNVSLSI 180
EQYEKIHEAY QILSKALKQA GLAPLNSKGE KLEAHVTTSK DQQTSSDQT TTTTSVIDTT 240
15 NDAQNLLTQA QTIVNTLKDY CPMLIAKSSS NGGTNGANTP SWQTAGGKKN SCATFGAEFS 300
AISDMISNAQ KIVQETQQLN ANQPKNITQP NNFNLNSPGS LTALAQSMLK NAQSQTEILK 360
LANQVASDFD KLSSGYLKDY IGKCDVSGVS SSNMTPQNMN TTWKGKGCAGV EETLTSLKAS 420
TTDFNNQTP QLDQAQTLAN TLTOELGNPN FKRVGIGSQ TNNGAMNGLG VQAGYKQFFG 480
QKRRWGLRYY GFFDYNHTYI KSSFFNSSSD VLTYGVSDDL LFNFLNDKNT NFLGKNNKIS 540
20 VGLFGGIALA GTSWLNSQFV NLKTISNVYS AKVNTANFQF LFNGLRLTNL ARPKKKDS DH 600
SAQHGMELGV KIPTINTNYY SYLGTKLEYR RLYSVYLYNV FAY 643
<212> Type : PRT
<211> Length : 643
SequenceName : SEQ ID 106
25 SequenceDescription :

Sequence

<213> OrganismName : Helicobacter pylori J99
30 <400> PreSequenceString :
MKKTILLSLM VSSLFAENDG VYMSVGYQIG EAAQMVKNTG EIQKVSNAEY NLNNLLTRYN 60
ELKQTASNTD SSTAQAIDNL EKSASRLKTT PNTANQAVSS ALSSAVGMWQ VIASNLANNS 120
LSSSEYEKLEK ATSQLLQNTL ENKNNNLKIE NDYDQLLTQA STIINTLQSQ CPGVDGGNGK 180
PWGINTSGNA CAIFGSTFNA INSMIDSAKK AAADARRTAP ESPNQONAFT NADFNKNLNQ 240
35 VSSVINDTIS YLKGDNLETI YNTIQKTPNS KGFQSLVSRs SYSYSLNETQ YSQFQTTTKE 300
FGHNPFPSVG LINSQSNGA MNGVGVQLGY KQFFGKKNKFF GIRYYGFFDY NYAYIKSNFF 360
NSASNVFTYG AGSDLLLNFI NGGSDRNRKV SFGIFGGIAL AGTTWLNNQS ANLKITNSAY 420
SAKINNTNFQ FLFNTGLRLQ GIHHGIELGV KIPTINTNYY SFMGAKLAYR RLYSLYLYNV 480
LAY 483
40 <212> Type : PRT
<211> Length : 483
SequenceName : SEQ ID 107
SequenceDescription :

45 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MPKASQVLFF GAFLSTSLQG FEAKLNGFVD QSSTIGFNQH KINKERGIYP MQQFATIAGY 60
50 LGLGFSLLPK KVSDHVLK GK IGMVGSIFY DGTKKFEDGS VAYNLFYGYD GFMGVYTNIL 120
QTD SLETQNM KHKNVRNYV FSDAYLEYAY KNYFEIKAGR YLSTMPYKSG QTQGFQVSGQ 180
YKHARLTWFS SWGRAFAYGS FLMDWFAART TYSGGFTKNN NGGYDSHGRK VLYGTHAVQL 240
TYKPHRFLIE GFYYLSPQIF NAPGVKIGWD SNPNFSGTGF RSDTAIGFF PIYYPWMIVK 300
SNGSPVYRYD TPATQNGQNL IIRQRFDINN YNVSIAFYKV FQNGANGWIGN MGNPSGVIMG 360
55 SNSVYAGFTG TALKRDAATI FLSCGGTHFA KKFTWKFATQ YNSVVSWEA RAMISLGYKF 420
TEYLSGSVDL AYYGVHTNKG FKPGENGVPV KNFPALYSDR SALYTALVAS F 471
<212> Type : PRT
<211> Length : 471
60 SequenceName : SEQ ID 108
SequenceDescription :

Sequence

65 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MLRLVSKTIC LSLISLFNPL EAFQKHQKDV FFVEAGFETG LLEGAQTKEQ AIAQNTQNTQ 60

5 KIYENPLTHP QTKEQPKEQN KSDTATPQSV YGRYYILQNT ILEKATELFT AANINGNGLT 120
FYSQNPVYVM AYNKDNAEFE GYGNNVSVVI QNFLPYNLNN IELSYTDAQG KAVNLGVIET 180
IPKDSQIILP ASLFNNFSND SPFNSDGLQQ LQTTTTPFSD ANTQSLFEKL SQITTNLQMT 240
YENTDPFSSG NNDPNGPLAS PKPHYECPGY KKSCQVASVS FTPQTAEELT NLMLDMIADV 300
DSKSWEEAVL NAPFQFSNSP SECGIDYPKC VNPFFNGLVD PKDEKYALTP EEVINSYRVA 360
NELTVNLLNA AKGFLGLGSQ LGSANAPDDD GFNQGVLGIA PFALDPEKLF GKNLNKVAI 420
ALRDIIHEYG HTLGYTHNGN MTYQVRVLCQ EGNGPEARCE GGHEVEKNGK EELEFSNGHE 480
VRDHDGYTYD VCSRFGGKNQ PAFPSNYPNS IYTNCAQVPA GLIGVTTAVW QQLINQNALP 540
INFANLNSQT SHLNAGLNAQ NFATSMVSAI AQNFSTTSTT TYRSSSKNFR SPILGVNVKI 600
10 GYQHYFNDYI GLAYYGIIQY NYAQANDEKI QQLSYGGGMD VLFDFTTYT NKKQDHPTKK 660
VFASSFGVFG GLRGLYSYY VFNQVKGSN LDIVTGFNYR YKHSKYSIGV SVPLIQSGIK 720
IASNNGIYAD SVVLNEGGSF FKVFFNYGWV F 751
<212> Type : PRT
<211> Length : 751
15 SequenceName : SEQ ID 109
SequenceDescription :

Sequence

20 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MQNFVFNKKW LIYSSLLPLF FLNPLMAEDD GFFMGVSYQT SLAVQRVDNS GLNASQDAST 60
YIRQNAIALE SAAVPLAYYL EAMGQQTRVL MQMLCPDPSK RCLLYAGGYQ NGQNNNGDTG 120
NNPPRGNVNA TFDMQSLVNN LNKLTQLIGE TLIRNPENLP NSKVFNVKFG NQSTVIALPE 180
25 GLANTMDALN NDITNALTTT WYNQTLTNKS FSTPSNTSVN FSPQVLQHLL QDGLATANNN 240
QTICSTQNCQ TATNEAKSIA QNAQNIFQAL MQAGILGGLA NEKQFGFTYN KAPNGSDSQQ 300
GYQSFSGPGY YTKNDNTTQA PLKALPAGAT IGSGNGQYTY HPSSAVYYLA DSIIANGITA 360
SMIFSGMQNF ANKAAKLIGT SSYNQMQDAI NYGESLLSNT VAYGDFITNW VAPYLDLNNK 420
GLNFLPNYGG QLNGANNQTP QLTPQQAQOE QKVIMNQLEQ ATNAPTPAQI NRILANPYSP 480
30 TAKTLMAYGL YRSKAVIGGV IDEMQTKVNQ VYQMGFARNF LEHNSNSNNM NGFGVKMGYK 540
QFFGKKRMFG LRYYGFDYDFG YAQFGTESSL VKATLSSYGA GTDFLYNVFT RKRGTETIDI 600
GFFAGIQLAG QTWKTNFLDQ VDG NHLKPKD TSFQFLFDLG IRTNFSKIAH QKRSRFSQGI 660
EFGLKIPVLY HTYYQSEGVV AKYRRDFSFY VGYNIGF 697
<212> Type : PRT
35 <211> Length : 697
SequenceName : SEQ ID 110
SequenceDescription :

Sequence

40 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTILLSLS LSLASSLLHA EDNGFFVSAG YQIGEAVQMV KNTGELKNLN EKYEQLSQYL 60
NQVASLKQSI QNANNIELVN SSLNYLKSFT NNNYNSTTQS PIFNAVQAVI TSVLGFWSLY 120
45 AGNYLTFFVV NKDTQKPASV QGNPPFSTIV QNCSGIENCA MNQTTYDKMK KLAEDLQAAQ 180
QNATTKANNL CALSGCATTO QGNPSSTVSN ALNLAQQLMD LIANTKTAMM WKNIVIAGVS 240
NVSGAIDSTG YPTQYAVFNN IKAMIPILQQ AVTLSQSNHT LSASLQAQAT GSQTNPKFAK 300
DIYAFQONQK QVISYAQDIF NLFSSIPKQD YRYLEKAYLK IPNAGKTPTN PYRQEVNLNQ 360
EIQTIONNVV YGGRVDAAL SVAKDVYNLK SNQTEIVTTY NNAKNLSQEI SKLPYNQVNT 420
50 KDIITLPHYDQ NAPAAGQYNY QINPEQQSNL SQALAAMSNN PFKKVGMISS QNNNGALNGL 480
GVQVGKQFF GESKRWGLRY YGFFDYNHGY IKSSFFNSSS DIWYGGGSD LLVNFINDSI 540
TRKNNKLSVG LFGGIQLAGT TWLNSQYMNL TAFNNPYSAK VNASNFQFLF NLGLRTNLAT 600
AKKKDSERSA QHGVELGIKI PTINTNYYSF LGTKLEYRRL YSVYLNIVFA Y 651
<212> Type : PRT
55 <211> Length : 651
SequenceName : SEQ ID 111
SequenceDescription :

Sequence

60 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MLKLASKTIC LSLISSFTAV EAFQKHQKDG FFIEAGFETG LLQGTQTQEQ TIATTQEKPK 60
65 PKPKPKPITP QSTYGYKYYIS QSTILKNATE LFAEDNITNL TFYSQNPVYV TAYNQESAEE 120
AGYGNNSLIM IQNFLPYNLN NIELSYTDDQ GNVVSLGVIE TIPKQSQIIL PASLFNDPQL 180
NADGFQQLQT NTTRFSDAST QNLFNKLSKV TTNLQMTYIN YNQFSSGNGS GSKPPCPPYE 240

NQANCVAKVP PFTSQDAKNL TNLMMLNMMAV FDSKSWEDAV LNAPFQFSDN NLSAPCYSDY 300
LTCVNPYNDG LVDPKLIASN KGDEYNIENG QTGSVILTPQ DVIYSYRVAN NIYVNLLPTR 360
GGDLGLGSQY GGPNGPGDDG TNFGALGILS PFLDPEILFG KELNKVAIMQ LRDIIEHYGH 420
TLGYTHNGNM TYQVRMCEE NNGPEERCQG GRIEQVDGKE VQVFDNGHEV RDTDGSTYDV 480
5 CSRFKDKPYT AGSYPNSTYT DCSQVPAGLI GVTSAVWQQL IDQNALPVDF TNLSSQTNYL 540
NASLNTQDFA TTMLSAISQS LSSSKSSATT YRTSKTSRPF GAPLLGVNLK MGYQKYFNDY 600
LGLSSYGIK YNYAQANNEK IQQLSYGVGM DVLFDFTNY TNEKNPKSNL TKKVFTSSLG 660
VFGGLRGLYN SYLLNQYKG SGNLNVGTGL NYRYKHSKYS IGISVPLVQL KSRIVSSDGA 720
YTNSITLNEG GSHFKVFFNY GWIF 744
10 <212> Type : PRT
<211> Length : 744
SequenceName : SEQ ID 112
SequenceDescription :
15 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
20 MRKLFIPLLL FSALEANEN GFFIEAGFET GLLEGTQTQE KRHTTTKNTY ATYNYLPTDT 60
ILKRAANLFT NAEAISKLKF SSLSPVRVLY MYNGQLTEN FLPYNLNNVK LSFTDAQGNV 120
IDLGVIETIP KHSKIVLPGE AFDSLKIDPY TLFLPKIEAT STSISDANTQ RVFETLNKIK 180
TNLVVNYRNE NKFKDHENHW EAFTPQTAE F'TNLMLNMIA VLDSQSWGDA ILNAPFEFTN 240
SPTDCDNDPS KCVNPGTNGL VNSKVDQKYV LNKQDIVNKF KNKADLDVIV LKDSGVVGLG 300
SDITPSNNDG GKHYGQLGVV ASALDPKKLF GDNLKTINLE DLRTILHEFS HTKGYGHNGN 360
25 MTYQVRPVTK DGQVEKDSNG KPKDSGGLPY NVCSLYGGSN QPAFPSNYPN SIYHNCADVP 420
AGFLGVTAAY WQQLINQNAL PINYANLGSQ TNYNLNASLN TQDLANSMLS TIQKTFVTSS 480
VTNHHSNAS QSFSPILGV NAKIGYQNYF NDFIGLAYYG IIKYNYAKAV NQKVQQLSYG 540
GGIDLLLDIF TTYSNKNSPT GIQTKRNFSS SFGIFGGLRG LYSYVVLNK VKGSGNLDVA 600
TGLNYRYKHS KYSVGISIP LQRKASVVSS GGDYTNSEVF NEGASHFKVF FNYGWVF 657
30 <212> Type : PRT
<211> Length : 657
SequenceName : SEQ ID 113
SequenceDescription :
35 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
40 MSLATSYNVS NNFSKFNIR VRGYLIQVLC NTPKMIQRGL NGVSFYGCSD YVNKGDCCKGV 60
LREINGSMKM VCLHCENTPI MEKVESGRGG AYACKNCNRK FYFIDLAKQN ERKKDLEKEK 120
KELLNKIEKQ KIKHLERFIL AGVKANIKEN SFFLGCKNYP KCEWTASMD S QDLKCPKCNR 180
LMKRKKNFKN NEFFTATSLT LNAIEFCLYI NLKKKETNV 219
<212> Type : PRT
45 <211> Length : 219
SequenceName : SEQ ID 114
SequenceDescription :
Sequence

50 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MEIKKYFLYA LFFLLFSGLF LSKLQAYKFN MSIVGVKSSY TKFGFNNQRY QPSKDIYPTG 60
SYTSLLGELN LSMGLYKGLR AEVGAMMAAL PYDSTAYQGN NIPNGQPGSR TDPFGAGIFW 120
55 QYIGWYAGHS GLNVQKPRLA MVHNAFLSYN YKKDKFSFGV KGGRYDAEEY DWFTSYTQGV 180
EGFVKYKDR LRVMYSDARA SASSDWFYF GRYTSGKAL MIADLKYEKD NLKINPYFYA 240
IFQRMYPAGI NITYDTNPNF NNGGFRFVGT FVGFFPIFAT PANQNDIILF QQVPLGKSGQ 300
TYFFRTRFY NKWQFGGSVY KNIGNANGDI GIYGDPLGYN IWTNSIYDAE INNIVGADVI 360
NGFLYVGSQY RGFSWKILGR WTDSPRADER SLALFLSYFS NKYNIRMDLK LEYYGNITKK 420
60 GYCIGYCGMY VPVDPNGPGT QPLTHNVYS RSHIMFNITY GFRIY 465
<212> Type : PRT
<211> Length : 465
SequenceName : SEQ ID 115
SequenceDescription :
65 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
5 MKKTILLSLS LSLASSLLHA EDNGFFVSAG YQIGEAVQMV KNTGELKLN LN DKYEQLSQSL 60
AQLASLK KSI QTANNIQAVN NALSDLKSFA SNNHTNKETS PIYN TAQAVI TSVLAFWSLY 120
AGNALS FHV T GLNDGSNSPL GRIHRDGNCT GLQQCFMSKE TYDKMKT LAE NLQKAQGNLC 180
ALSECSSNQ S NGGKTSMTTA LQTAQQLMDL IEQTKVSMVW KNIVIAGVTN KPNGAGAIT S 240
TGHVTDYAV F NNIKAML PIL QQALTLSQSN HTLSTQLQAR AMGSQTNREF AKDIYAL AQN 300
QKQILSNASS IFNLFNSIPK DQLKYLENAY LKVP HLGKTP TNPYRQNVNL NKEINAVQDN 360
VANYGNRLDS ALSVAKDVYN LKSNQTEIVT TYNDAKNLSE EISKLPYNQV NVTNIVMSPK 420
10 DSTAGQYQIN PEQQSNLNQA LAAMSN NP FK KVGMISSQNN NGALNGLGVQ VGYKQFFGES 480
KRWGLRYYGF FDYNHGYIKS SFNSSSDIW TYGGGSDLLV NFINDSITRK NNKLSVGLFG 540
GIQLAGTTWL NSQYMNLTAF NNPYS AKVNA SNFQFLFNLG LRTNLATAKK KDSERSAQHG 600
VELGIKIPTI NTNYY SFLGT KLEYRRLYSV YLNYVFAY 638
<212> Type : PRT
15 <211> Length : 638
SequenceName : SEQ ID 116
SequenceDescription :

Sequence
20 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
25 MKLKKRKVAA TLLKRLTLPL LFTTGSLGAV TYEVHGDFIN FSKVGFNRSP INPVKGIYPT 60
ETTFVNLTGKL EGSVHLGRGW TVNVGGVLGG QVYDNTRYDR WAKDFTPPSY WDKTSCGTDS 120
LSLCMNATKM WQQQGP GGII DPRGIGYMYM GEWNGLF PNY YPANAYLP GH SRRYEVYKAN 180
LTYDSDRVHM VMGRFDVTEQ EQMDWIYQLF QGFYGT FKL T KNMKFLLFSS WGRGIADGQW 240
LFPIYREKPW GIHKAGIYR PTKNLMIH PY VYLIPMVGT L PGAKIEYDTN PEFSGRGIRN 300
KTTFYVLYDY RWNNAEYGRY APARYNTWDP FLDNGKWRGL QGPGGATLYL HHHIDINNYF 360
VVG GAYLNIG NPNMNLGTWG NPVALDGIEQ WVGGIYSLGF AGIDNITDAD AFTEYVKG GG 420
30 KHGKFSWSVY QRFTTAPRAL EYGIGMYLDY QFSKHVKAGL KLVWLEFQIR AGYNPGTGFL 480
GPNGQPLNLN NGLFESSAFA QGPQNMGGIA KSITQDRSHL MTHISYSF 528
<212> Type : PRT
<211> Length : 528
SequenceName : SEQ ID 117
35 SequenceDescription :

Sequence

<213> OrganismName : Helicobacter pylori J99
40 <400> PreSequenceString :
MKNFSPLYCL KKLKKRH LIA LSLPLLSYAN GFKIQEQSLN GTALGSAYVA GARGADASFY 60
NPANMGFTND WGENRSEFEM TTTVINIPAF SFKVPTTNQG LYSVTSLEID KSQQNILGII 120
NTIGLGNILK ALGNTAATNG LSQAINRVQG LMNLTN QKV TLASKPDTQI VNGWTGTTNF 180
VLPKFFYKTR THNGFTFGGS FTAPSGLG MK WNGKGGEFLH DVFIMMVELA PSMSYTINKR 240
45 FSVGVGLRGL YATGSFNNTV YVPLEGASVL SAEQILNLPN NVFADQVPSN MMTLLGNIGY 300
QPALNCQKAG GDMSDQSCQE FYNGLKKIMG YSGLIKASAN LYGT TQVVQK SNGQGVSGGY 360
RVGSSLRVFD HGMFSVVYNS SVTFNMKGGL VAITELGPSL GSVLTKGSLN INVSLPQTLS 420
LAYAHQFFKD RLRVEGVFER TFWSQGNKFL VTPDFANATY KGLSGTVASL DSETLKKMVG 480
LANFKSVMNM GAGWRDTNTF RLGVTYMGKS LRLMGAIDYD QAPSPQDAIG IPDSNGYTVA 540
50 FGTKYNFRGF DLGVAGSFTF KSNRSSLYQS PTIGQLRIFS ASLGYRW 587
<212> Type : PRT
<211> Length : 587
SequenceName : SEQ ID 118
SequenceDescription :

Sequence
55 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
60 MAFQVNTNIN AMNAHVQSAL TQNALKTSLE RLSSGLRINK AADDASGMTV ADSLR SQASS 60
LGQAIANTND GMGIIQVADK AMDEQLKILD TVKV KATQAA QDGOTTESRK AIQSDIVRLI 120
QGLDNIGNTT TYNGQALLSG QFTNKEFQVG AYSNQSIKAS IGSTTS DKIG QVRIATGALI 180
TASGDISLTF KQVDGVNDVT LESVKVSSSA GTGIGVLA EV INKNSNRTGV KAYASVITTS 240
DVAVQSGSLS NLTLNGIHLG NIADIKKND S DGRLVA AINA VTSETGVEAY TDQKGRNLNR 300
65 SIDGRGIEIK TDSVSNGPSA LTMVNGGQDL TKGSTNYGRL SLTRLDAKSI NVVSASDSQH 360
LGFTAIGFGE SQVAETTVNL RDVTGNFNAN VKSASGANYN AVIASGNQSL GSGVTTLRGA 420
MVVIDIAESA MKMLDKVRSD LGSVQNQMIS TVNNISITQV NVKAAESQIR DVDFAEESAN 480

FNKNNILAQS GSYAMSQANT VQQNILRLLT 510
<212> Type : PRT
<211> Length : 510
SequenceName : SEQ ID 119
5 SequenceDescription :

Sequence

10 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MAGTQAIYES SSAGFLSQVS SIISSTSGVA GPFAGIVAGA MTAALPIVV GFTNPQMTAI 60
MTQYNQSIAS AVSVPKKAAN QQYNQLYQGF NDQSMAGVNN ILNISKLTGE FNAQGNTQSA 120
QISAVNSQIA SILASNTTPK NPSAIEAYAT NQIAVPSVPT TVEMMSGILG NITSAAPKYA 180
LALQEQLRSQ ASNSSMNDTA DSLDSCTALG ALVGSSKVFF SCMQISMTPM SVSMPTVYAK 240
15 YQAVATKALT SGVNPMTTPA CFIGDKVLAV YCYAEKVAEI LREYYIEFVK NNTNLLQNAS 300
QMILNQSGLA TSTYDTQAIS NISSLYNYNI VANKSFLKSH LTYLDYIKDK LKGQKDSYLT 360
ERVQTKIIVK 370
<212> Type : PRT
<211> Length : 370
20 SequenceName : SEQ ID 120
SequenceDescription :

Sequence

25 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MTNEAINQQP QTEAAFNPQQ FINNLQVAFI KVDNVVASFD PNQKPIVDKN DRDNRQAFEK 60
ISQLREEFAN KAIKNPTKKK QYFSSFISKS NDLIDKDNLI DTGSSIKSFQ KFGTQRYQIF 120
MNWVSHQNDP SKINTQKIRG FMENIIQPP I SDDKEKAEFL RSAKQAFAGI IIGNQIRSDQ 180
30 KFMGVFDES LKERQAEKNG EPNGDPTGGD WLDIFLSFVF NKKQSSDLKE TLNQEPVPHV 240
QPDVATTTTD IQSLPPEARL LDERGNFSK FTLGDMNMLD VEGVADIDPN YKFNQLLIHN 300
NALSSVLMGS HNGIEPEKVS LLYGNNGGPE ARHDWNATVG YKNQRGDNVA TLINVHMKNG 360
SGLVIAGGEK GINNPSFYLY KEDQLTGSQR ALSQEEIQNK VDFMEFLAQN NAKLDNLSKK 420
EKEKFQNEIE DFQKDSKAYL DALGNDHIAF VSKKDKKHLA LVAEFGNGEL SYTLKDYGKK 480
35 ADKALDREAK TTLQGS LKHD GVMFVDYSNF KYTNASKSPD KGVGATNGVS HLEAGFSKVA 540
VFNLPLNNL AITSVVRQDL EDKLIAGLS PQEANKLVKD FLSSNKELVG KALNFKAVA 600
EAKNTGNYDE VKQAQKDLEK SLKKRERLEK DVAKNLESKS GNKNKMEAKS QANSQKDEIF 660
ALINKEANRD ARAIAYAQNL KGIKRELSKD LENINKDLKD FSKSFDEFKN GKNKDFSKAE 720
ETLKALKGSV KDLGINPEWI SKVENLNAAL NEFKNGKNKD FSKVTQAKSD LENSIDVVI 780
40 NQKITDKVDN LNQAVSVAKA TGDFSGVEQA LADLNKFSKE QLAQQAQKNE DFNTGKNSAL 840
YQSVKNGVNG TLVGNGLSKA EATTLNKNFS DIKKELNAKL GNFNNNNNNG LENSTEPIYT 900
QVAKKVKAKI DRLDQIASGL GDVGQAASFL LKRHDKVDDL SKVGLSANHE PIYATIDDLG 960
GPFPLKRHDK VDDLKSVGLS REQKLTQKID NLNQAVSEAK ASHFDNLDQM IDKLKDSTKK 1020
NVVNLYVESA KKVPTSLSAK LDNYATNSHT RINSNVKNGT INEKATGMLT QKNSEWLKLV 1080
45 NDKIVAHNVG SAPLSAYDKI GFNQKNMKDY SDSFKFSTR L SNAVKDIKSG FVQFLTNI FS 1140
MGSYS LMKAS VEHGVKNTNT KGGFQKS 1167
<212> Type : PRT
<211> Length : 1167
50 SequenceName : SEQ ID 121
SequenceDescription :

Sequence

55 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKTNGHFKDF AWKKCFLGAS VVALLVGCSP HIIETNEVAL KLNYPHASEK VQALDEKILL 60
LRPAFQYSDN IAKYENKFK NQTTLVKEEI LQNGQYKVIN VDSSDKDDFS FAQKKEGYLA 120
VAMNGEIVLR PDPKRTIQKK SEPGLLFSTG LDKMEGVLP AGFVKVTILE PMSGESLDSF 180
TMDLSELDIQ EKFLKTTHSS HSGGLVSTMV KGTDNSNDAL KSALNKIFAS IMQEMDKKLT 240
60 QRNLESYQKD AKELKNKRN 260
<212> Type : PRT
<211> Length : 260
SequenceName : SEQ ID 122
SequenceDescription :

65 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
5 MKSKLKLKRY LLFLPLLPLG TSLANTYLL QDHNTLTPYT PFTTPLNGGL DVVRAAHLHP 60
SYELVDWKRV GDTKLVALVR SALVRVKFQD TTSSDQSNTN QNALSFDTOE SQKALNGSQS 120
GSSDTSGSNS QDFASYVLIF KAAPRATWVF ERKIKLALPY VKQESQSGSD QGSNGKGSly 180
KTLQDLLVEQ PVTPTYTPNAG LARVNGVAQD TVHFGSGQES SWNSQRSQKG LKNNPGPKAV 240
TGFKLDKGRA YRKLNESWPV YEPLDSTKEG KGKDESSWKN SEKTTAENDA PLVGMVGSga 300
AGSASSLQGN GSNSSGLKSL LRSAPVSVPP SSTSNQTLSTL SNPAPVGPQA VVSQPAGGAT 360
AAVSVNRTAS DTATFSKYLN TAQALHQMVG IVPGLEKWGG NNGTGCVVASR QDATSTNLPH 420
10 AAGASQTGLG TGSPREPALT ATSQRAVTVV AGPLRAGNSS ETDALPNVIT QLYHTSTAQL 480
AYLNGQIVVM GSDRVPSLWY WVVGEDQESG KATWWAKTEL NWGTDKQKQF VENQLGFKDD 540
SNSDSKNSNL KAQGLTQPAY LIAGLDVVAD HLVFAAFKAG AVGYDMTTDS SASTYNQALA 600
WSTTAGLDSD GGYKALVENT AGLNGPINGL FTLLDTFAYV TPVSGMKGGS QNNEEVQTTY 660
FVKSDQKATA KIASLINASP LNSYGDDGVT VFDALGLNFN FKLNEERLPS RTDQLLVYGI 720
15 VNESELKSAR ENAQSTSDDN SNTKVKWTNT ASHYLPVPYY YSANFPEAGN RRRAEQRNGV 780
KISTLESQAT DGFANSLNLF GTGLKAGVDP APVARGHKPN YSAVLLVRGG VVRLNFPNPD 840
DKLLDSTDKN SEPISFSYTP FGSAESAIDL TTLKDVITYIA ESGLWFYTFD NGEKPTYDGK 900
QQQVKNRKGy AVITVSRGTI EFNEDANTTT LSQAPAALAV QNGIASSQDD LTGILPLSDE 960
FSAVITKDQT WTGKVDIYKN TNGLFKEDDQ LSENVKRRDN GLVPIYNEGI VDIWGRVDFA 1020
20 ANSVLQARNL TDKTVDEVIN NPDILQSFfk FTPAFDNQRA MLVGEKTSdT TLTVKPKIEY 1080
LDGNFYGEDS KIAGIPLNID FPSRIFAGFA ALPSWVIPVS VGSSVGILLI LLILGLGIGI 1140
PMYKVRKLQD SSFVDVFKKV DTLTTAVGSV YKKIITQTSV IKKAPSALKA ANNAAPKAPV 1200
KPAAPTAPRP PVQPPKKA 1218
<212> Type : PRT
25 <211> Length : 1218
SequenceName : SEQ ID 123
SequenceDescription :

Sequence
30 -----
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
35 MHQTKKTALS KSTWILILTA TASLATGLTV VGHFTSTTTT LKRQQFSYTR PDEVALRHTN 60
AINPRLTPWT YRNTSFSSLP LTGENPGAWA LVRDNSAKGI TAGSGSQOTT YDPTRTEAAL 120
TASTTFALRR YDLAALYD LDFSKLNPQT PTRDQTGQIT FNPFGGFLS GAAPQQWNEV 180
KNKVPVEVAQ DPSNPYRFAV LLVPRSVVYY EQLQRGLGLP QORTESGQNT STTGAMFGLK 240
VKNAEADTAK SNEKLQGAEA TGSSTTSGSG QSTQRGSSG DTKVKALKIE VKKKSDSEDN 300
GQLQLEKNDL ANAPIKRSEE SGQSVQLKAD DFGTALSSSG SGGNSNPGSP TPWRPWLATE 360
QIHKDLPKWS ASILILYDAP YARNRTAIDR VDHLDPKAMT ANYPPSWRTP KWNHHGLWDW 420
40 KARDVLLQTT GFFNPRRHPE WFDGGQTVAD NEKTGFDVDN SENTKQGFQK EADSDKSAPI 480
ALPFEAYFAN IGNLTWFGQA LLVFGGNGHV TKSHTAPLS IGVFRVRYNA TGTSATVTGW 540
PYALLFSGMV NKQTDGLKDL PFNNNRWFEY VPRMAVAGAK FVGRELVLG TITMGDTATV 600
PRLLYDELES NLNLVAQGGG LLREDLQLFT PYGWANRPDL PIGAWSSSSS SSHNAPYYFH 660
NNPDWQDRPI QNVVDAFIKP WEDKNGKDDA KYIYPYRYSg MWAQVYNWS NKLTDQPLSA 720
45 DFNENAYQP NSLFAAILNP ELLAALPDKV KYGKENEFAA NEYERFNQKL TVAPTQGTNW 780
SHFSPTLSRF STGFNLVGSV LDQVLDYVPW IINGYRYGNN HRGVDDITAP QTSAGSSSGI 840
STNTSGRSRF LPTFSNIGVG LKANVQATLG GSQTMITGGS PRRTLQANL QLWTGAGWRN 900
DKASSGQSDE NHTKFTSATG MDQQGQSGTS AGNPDSLQKD NISKSGDSL TQDGNADQQ 960
EATNYTNLPP NLTPTADWPN ALSFTNKNNA QRAQLFLRGL LGSIPVLVNR SGSDSNKFQA 1020
50 TDQKWSYTDL HSDQTKLNLP AYGEVNGLLN PALVETYFGN TRAGGSGSNT TSSPGIGFKI 1080
PEQNNDKAT LITPGLAWTP QDVGNLVVSG TTVSFQLGGW LVTFTDFVKP RAGYLGLQLT 1140
GLDASDATQR ALIWAPRPWA AFRGSWVNL GRVESVWDLK GVWADQAQSD SQGSTTTATR 1200
NALPEHPNAL AFQVSVVEAS AYKPNTSSGQ TQSTNSSPYL HLKPKKVTQ SDKLDDDLKN 1260
LLDPNQVRTK LRQSFQTDHS TQPQPQSLKT TTPVFGTSSG NLSSVLSGGG AGGGSSSGSQ 1320
55 SGVDLSPVEK VSGWLVGQLP STSDGNTSST NNLAPNTNTG NDVVGVGRLS ESNAKMND 1380
VDGIVRTPLA ELLDGEQTA DTGPQSVKFK SPDQIDFNRL FTHPVTDLFD PVTMLVYDQY 1440
IPLFIDIPAS VNPKMVRLKV LSFDTEQSL GLRLEFFKPD QDTQPNNNVQ VNPNGDFLP 1500
LLTASSQGPQ TLFSPFNQWP DYVLPLAITV PIVVIVLSVT LGLAIGIPMH KNKQALKAGF 1560
ALSNQKVDVL TKAVGSVFKE IINRTGISQA PKRLKQTSAA KPGAPRPPV PKPGAPKPPV 1620
60 QPPKKPA 1627
<212> Type : PRT
<211> Length : 1627
SequenceName : SEQ ID 124
SequenceDescription :

Sequence
65 -----

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MGYKLKRWPL VAFTEFTGIGL GVVLAACSAL NTSNLFPRQN RSKQLIGFTE NNIIEKPEAVL 60
KAALAEDNGT ETILRVNFGE ALKSWYQNNK DRNIATRLTI FSENVDEHD NLLDQKQQA 120
5 PINWPIELQK EYDQWGGSES SWKALKLYDR LIADFQSLIF SNIVANVQLT DGSDQFKPTT 180
KDNLDSTSNK IKFVNSKPND PNGEFFANLQ AYLFAQWVVE ENPLPLTQAF FAYQAPKDGL 240
DSLYDQAAIG SALQLGYAFP AFREPNNGQS QGKTTFDPTP NSAQNFGDFI KAVFPEQKNG 300
QTQQSNTSSR TGLFDWQTKW NTNGAANKLL VTKSNLRGAF KGVGLATAII DQYEYLVGGS 360
KTSSLPEVKV DSNKSNQNP L DSFFMEGKDA VAIRSIVSRA KIAMTDQTPG FKVNPAPVKV 420
10 KQSQQNDTFY QNQRKLSGGQ SGDNN SQGKH HYLQDAVRLT SSQAMAAAST GADSSSGTNV 480
GGSSGGNSVL IPLPRSAALT HTQQQVQQTT STLQTPVYAR GDDGTALAI DGGDYFLANN 540
KRDFTKQADI LLYRYLQAKS NNFKENGVEF SLNLLESGL FQTTWAQTGLT AKLYGALVAM 600
MSGGQGTQVK GSVQGSSRAA SVSVQTTQON RQSTDTQES EVVKLAKSL KSSADLAKPF 660
TDNPTFKKAL TDIQSEYKDY LAAAGKLSEF KKDLEGEVSL QQAIIDRADK YIQLEKQAQK 720
15 SAIGLQGPLP YQRASDGSYP ALEKFFIPED SAADGKVKAS ESGSALVTL KTTDSQKSTN 780
TVKQPDIKPT RENNDKKLKQ LTSDVETKAS SLITKWGATP QIGSQFSEIV SLKSKDNKPQ 840
TNMILALLSD VGIKWTILN SFKEWFTNT NDFKNNYDSE KKEKLGNEYK DFNDLVKQTL 900
YLRWQRLTS KEKFGYYKEL GSVKAQAQAS GMVSLSSSAA VANAVASSGM QKSGDQTLLE 960
LGKKAFESL EASSSDGQYK YLRLSTLMW LVKDGAKNYK RLLQQAITVG TRAFVSWTVS 1020
20 YDDTATASAA AAKAQVAVLK TAQATNTQSD NPFNKVQNP DYVQGETNW FNDKSTPIK 1080
DSLLESESTY NFTAEPFDDK TKSQKRSTGG TTNEKHFFGF NGLTINSPOS VSTASAGLTE 1140
QIFNNFGQLV TSSDKSGALS QYKDKATLKR LIQNTNSDAE LNAFGEVLHR AVNVDTSNLG 1200
RFNSSGEPLI SFDNKKKFLV DVVDKLDVY FNFEGYVGQ TKVKMSDSSS SSQGTKTIRK 1260
PKPHHSPRTR VSRLWAMSFR LPTRTLTKFL LVEKLIRTVL 1300
25 <212> Type : PRT
<211> Length : 1300
SequenceName : SEQ ID 125
SequenceDescription :
30 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MKKLLIKPQF WFLTLGGFIS SSVILVACAT PSNSALQTVF KARSNQFFNG EQGSLQNALA 60
35 TALKDPEANK QFVAAPLLKA LTAWYENNQD KQVTQFFKDT KKSVDQYNQ AVDKVVSASR 120
NKNLFVQQDL LDSAGGVRNL KSPEVWVTAH 150
<212> Type : PRT
<211> Length : 150
SequenceName : SEQ ID 126
40 SequenceDescription :
Sequence

<213> OrganismName : Mycoplasma pneumoniae
45 <400> PreSequenceString :
MQQQGETKDQ YNTFGLRLVR NSVGVSVLGL DGFVKFIKGG SGGSNNGSSS AKKIDKEEQK 60
KFLKFRAFQA KIGTFYNTNF AFSFPLNETL KGWFDKHRGL ILANALVKVT LDTKEKASKA 120
LVDAFSSYKN WLSEYTPVGL ATTMSIFYFD QMKALNNKLL ERVRSLNQNV NQANPTPWLN 180
GLSAKLPHYVN TNGNYEKLNN YFTFLITKVL WPKVGTEDTN VSEEKSKLKT KTEDVNKIRE 240
50 KILNNIDSKL KTFVQKLKPT LAPRPAYSNV ILLNINNDKV WSAGANWSLA VLLDPKKVNP 300
LSFMLLKQMF DQNSLFKKAK TLFENIQNKA KTSGSGKSGT TTNDADALS KVIGNYYYNT 360
WAKLTDKSIY GNLKDDKFDD LFKLAFDSSI NEKSFNVYD AVIEHYRFIY TLEWLVDKNL 420
KNFKDLLKAN LKFGEIAFIA YKNTETQNF NPQGIFGSYF NYENETNAAK SATQIIDPNS 480
PFYKTTTKPE AKTTQSANTA VMVQNTQMNN QQTNSYGFTG LSTSSGSM LG AATQQAILDQ 540
55 ITKTSLQQYG SQADLKKIIG ETKNQLLLDR IANQLIALKP NTSGNSGTQK TIAAYFQTD 600
VGNPTLDFKA KQKLLLDVLD QYKDFFGNNA QAVQRDSGKS GTGNYLTYTD GSKITYLQF 660
SYKDIDGLSL SSSNGTSSKF ASDVVAALLL FQAAYKGTQQ LALSSINKPQ LPIGDKRIKT 720
GIDLLK 726
<212> Type : PRT
60 <211> Length : 726
SequenceName : SEQ ID 127
SequenceDescription :
Sequence

65 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

5 MKSFLRKPKF WLLLLGGLST SSIILSACAT PSNSALQAVF KPTSNQFFNG EHGTIQSALN 60
TALRDPETNK KFVAAPLLKA LEAWYENNQD KNITQFLKDT KTNVDNQYKT VVDKVVSAAPR 120
NKSFLVQQDL LDSSGGSEAT WKARKLFEQL ISDFASRVFQ KNYLSYKENG KVSAGPFLYD 180
TISKNSNWQN IVFDAVNFPE TNDDFFAKIQ SEVFDQWAEY TDPTIISSVT LKYSAPN 237

10 <212> Type : PRT
 <211> Length : 237
 SequenceName : SEQ ID 128
 SequenceDescription :

15 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :

20 MINFLFNQMN ALNNKFLERA KALNQNVNQA NPTPWLNGLS AKLPYVRTNG NYEKLNNYFT 60
FLIVKYMWKK VGNEDASLSK DSSINKLTKT TEDVNKIRDK ILEDIQKKVQ EFVKNKLKPT 120
LAPRQYTSNV ILLNVNNDKV WSMGANWALA NLLDTSKINP LSFMLLKQTF DQNDLFKKAK 180
KLFEDIQSKT NGGSSGGMQG SNTSSSEGAD ALSKVIGNYY YNSWAKLTDK SIYGNPKDNK 240
FDDLFLKAFE DSINEKSFNV DYKAVIEHYR FIYTLLEWLVN GNLKNFKDLL KANLKFGEIA 300
25 FIAYKNTETK EFSNPQGVFG SAFNYENETN EVKIAAQNLD PNNFFYKTTT KPEEVKTAQN 360
GASMMVMQQK MQSTMQDSNH YGFTGLNTST SSMLGAATQQ AILDQITKNS LQQYGSQQEL 420
KTLIEKTNNQ LLLDRIASQL SGLNPSTTGN SNNKGKGNIA TYFQLDAIGN PTLSFQQKRK 480
LLLDVLDQYK DFFGTNTQAA QRDSGKGGHG SYSTYQDGS D KITYLQFSYK DIDNLSLSDK 540
GNSKLASDVV AALLLFQAAD KGTQQLALSA IN 572

30 <212> Type : PRT
 <211> Length : 572
 SequenceName : SEQ ID 129
 SequenceDescription :

35 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :

40 MKKFLRKPKF WLLTLGGFLS TSVILAACAT PSNSALQTVF KARSSQFFNG EQGSLQSALT 60
TALKNPVANK QFIAAPLLKA LEAWYENNED KKITQFLKDT KSNVDSQYTT AVDKVVSASR 120
NKSFLVQQDL LDNAGGSEAT WKAQKLLEQL ISDFASRVFQ KNYLNYKKDG QVSTGPFTYD 180
ELHKEESWKN FEFSAPRFSE TNDDFFAKIQ SQVFDQWVEY TDPTLISQVN YKYSAPSQGL 240
GQIYNREKLK DKLTPSYAFP FFAEEKDIAP NQNVGNKRWK QLVKGEGAIT DNNIGQSGTN 300
SQKTGLLK YR NESNKGDFLD FPLNLSDTNE TKQLVDASNI VDQLEAANLG AALNLKLQVF 360
45 EQDNDELPIQI KELKEDLNNT IVVDKSKDVE KASKTNALFY NDQEGKQQQS DSDPIAGALD 420
DIFAQNTSEG TNLSKLAEQV KKAATKMEA KTAVLRTNNS KGQQNNYVVL DAAIPTFNST 480
TSKSKNNSAS NEVLVALKSG SINLRQVQQT DQNSYSPIKF RIVRNSTGVT VFGLDGGSY 540
LKQDSTNKKS VSKQSLTLLT KSSSGNSNKV LRDLDKQKQF LKFRFQAKT NTFYSTNFAF 600
SFPLNETLKS WFDKHRELIL ANALVNASLD QKDKASKALT EAFNPYKELI KEFAPVALAT 660
50 TMISFYFDQM KALNNKLLER ARNLNQNVNQ ANPTPWLNGL SAKLPYVNTN GNYEKLNNYF 720
TFLITKTLWP KVGQEETSIS EESNKLKTKT ADVDKIRDKI LENIQTKVND FVKNKLKPAL 780
APRPAYSNI LLNVNNDKVL SSGANWSLAS LLQSDKVNPL SFMLLKQAFD NNDLFKKAQK 840
LFKDIQEKSS NNGGMQSSST TNSDADALSK VIGNYYTWTW AKLTDKSIYG NPKDNKFDEL 900
FKLAFEASID EKSFNVDYKA VIDHYRFIYT LQWLVDQKLG NFKSLLKTNL KFGEVAFIAY 960
55 KNTETTNFSN PQGVFGSYFN YENSASEVKE STQTLDPNNF FYKTTTKPTV QAIQQVASLA 1020
LVQKQQMQQN STDHYGFTGL STSTSSMFDA SSRDAILQQI TKTSLQQYGS KDQLKKIIQG 1080
TNNQLLLDRI AVQLSGLNPS TTNGGSGKTI ATYFQVDAVG NPTLDFQAKR KLLLDLLDQY 1140
QNYFGNGAQK SQRDSTPSGT GNYLTYQNGS DKYTYTQFTY QDIDSLSLTT TSGTNNKIAS 1200
DVVAALLLFQ AADKGTQQLA LSAINKPQLN IGDKRIESGL KLLK 1244

60 <212> Type : PRT
 <211> Length : 1244
 SequenceName : SEQ ID 130
 SequenceDescription :

65 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :

70 MVGSGAAGSA SSLQNGSNS SGLKSLRSA PVSVPSSSTS NQTLSSLNPA PVGPQAVVSQ 60
PAGGATAAVS VNRTASDTAT FSKYLNATAQ LHQMGVIVPG LEKWGGNNGT GVVASRRDAT 120
STNLPHAAGA SQTGLGTGSP REPALTATSQ RAVTVVAGPL RAGNSSETDA LPNVITQLYH 180
TSTAQLAYLN GQIVVMSSAR VPSLWYWVG EDQESGKATW WAKTELNWGT DKQKQFVENQ 240

LGFKDDSNDS SKNSNLKTQG LTQPAYLIAG LDVVADHLVF AAFKAGAVGY DMTTDSNAST 300
YNQALVWSTT AGLDSDGGTR LW 322
<212> Type : PRT
<211> Length : 322

5 SequenceName : SEQ ID 131
 SequenceDescription :

Sequence

10 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :
MPVFLKLTHT IRKVLRVARL SRLALLSLTA VIFSGCANIN LISAVGSSSV QPLL SKLSSH 60
YVLNHNDKDN LVEISVQAGG SSAGVKAITK GLADIGNVSK NTKSYAEENK QLWMDKKLKT 120
ITLGKDAIAV IYKAPSEFKG KLVLT KDNLN DLYDLFAGSK SVDINKFVEN GQTTKNSNHN 180
15 LIGFPRTGGA FASGTAEAF LKFSGLTQTKT LDKDSKEILE GQRNYGPNAR PTSETNIEAF 240
NTFVTTLRQP NLYGMVYLSL GFVNNNMNLI KSEGFEVLKV KYDNNAVTPS SQAVSSNTYK 300
WVRPLNSVVS LLPKQKTLPS IQRFFNWLLF SNNSEIKKIY DDFGVLELTA DEKKKMFKTG 360
NAEMSNIANF WDDYSLNNQ TFGAL 385
<212> Type : PRT
20 <211> Length : 385
 SequenceName : SEQ ID 132
 SequenceDescription :

Sequence

25 <213> OrganismName : Mycobacterium tuberculosis H37Rv
 <400> PreSequenceString :
MSFAVLPPET NSARLYVGAG LAPMLDAAAA WDGLADELGS AAASFSAVTA GLAGSSWLGA 60
ASTAMTGAAA PYLGWLSAAA AQAQQAATQT RLAAAAFEAA LAATVHPAII SANRALFVSL 120
30 VVSNNLGQNA PAIAATEAAY EQMWAQDVAA MFGYHAGASA AVSALTPFGQ ALPTVAGGGA 180
LVSAAAAQVT TRVFRNLGLA NVGEGNVGNG NVGNFNLGSA NIGNGNIGSG NIGSSNIGFG 240
NVGPGLTAAL NNIGFGNTGS NNIGFGNTGS NNIGFGNTGD GNRGIGLTGS GLLGFGGLNS 300
GTGNIGLFNS GTGNVGIGNS GTGNWGIGNS GNSYNTGFCN SGDANTGFFN SGIANTGVGN 360
AGNYNTGSYN PGNSNTGGFN MGQYNTGYLN SGNYNTGLAN SGNVNTGAFI TGNFNNGFLW 420
35 RGD HQGLIFG SPGFFNSTSA PSSGFFNSGA GSASGFLNSG ANNSGFFNSS SGAIGNSGLA 480
NAGVLVSGVI NSGNTVSGLF NMSLVAITTP ALISGFFNTG SNMSGFFGGP PVFNLGLANR 540
GVVNILGNAN IGYNINILGSG NVGDFNIGLS GNLGSQNILG SGNVGSFNIG SGNIGVFNVG 600
SGSLGNYNIG SGNLGIYNIG FGNVGDYNVG FGNAGDFNQG FANTGNNNIG FANTGNNNIG 660
IGLSGDNQQG FNIASGWNSG TGNSGLFNSG TNNVGIFNAG TGNVGIANSN TGNWGIGNPG 720
40 TDNTGILNAG SYNTGILNAG DFNTGFYNTG SYNTGGFNVG NTNTGNFNVG DTNTGSYNPG 780
DTNTGFFNPG NVNTGAFDTG DFNNGFLVAG DNQQQIAIDL SVTTPFIPIN EQMVIDVHNV 840
MTFGGNMITV TEASTVFPQT FYLSGLFFFG PVNLSASTLT VPTITLTIGG PTVTVPI SIV 900
GALESRTITF LKIDPAPGIG NSTTNPSSGF FNSGTGGTSG FQNVGGGSSG VWSGLSSAI 960
GNSGFQNLGS LQSGWANLGN SVSGFFNTST VNLSTPANVS GLNNIGTNLS GVFRGPTGTI 1020
45 FNAGLANLQ LNIGSANLGD FNLGSGNVGS FNVFSGNQGS YNIGPANLGN YNIGFANLGN 1080
YNIGFGNAGD FNQGFANTGN NNIGFANTGN NNIGIGLSGD NQQGFNFAGG WNSGTANIGL 1140
FNSGTNNVGI GNSGTGNWGI GNSGSGNTGI GNTGSTNTGF FNTGIVNTGV ANAGSYNTGW 1200
YNTGDTNTGI ANLGDFNTGF YNTGNFSTGF ANQGDIA TGA FITGDMNGA FWRGDQQGLF 1260
SAGYRVHVPE IPAHTVVEVP VNIPITASFT NTVYSGITLE QINFGFTIDI AGIPLLAGAI 1320
50 SKAVLPPITG TGPAITVNIG DPGGSTAIRI PATASVGPF D VTFVNIAATT GFFNATTDPS 1380
SGFFNGGPGT VSGIANIGAN ISGFQNVANS ATSGFNNGYS LQSGLANLGD TVSGVFNTGI 1440
GAPANVSGMF NIGSNLAGFF HDQATGMSMF NLGLGNIGQF NVGFSNVGDS NAGLANIGSF 1500
NLGSGNLGSF NVFSGNQGSY NIGPANLGN YNIGLNLGSY NFGFGNAGDF NLGFANTGNN 1560
NIGFANTGNN NIGIGLSGDN QQGFNFAGGW NSGSGNSGLF NSGTNNIGLF NSGTGNIGIG 1620
55 NSGTGNWGIA NTGDTNTGIF NTG DVNTGLL NAGNVNTGIF NTGHYNTGSF NAGSFNTAGF 1680
NPGSYNTGYL NTGSYNTGLA NSGDVNTGGF ITGNYSNGFW WRGDYQGLAG ISQTITVPDT 1740
AVPVKLHVPI FLDIPVTGTL GTFTVHGFRF PEITGDIFLI GIPFNAATLD AFSFPNISIV 1800
LPNIGINLGS GPDPLIDIAG TGGLLPKIP LIDIPAAPGF GNSTTPSSG FFNAGTGTVS 1860
GVGNVGSNSS GFFNLTS GSS GISGVQNFGE LISGGFNFGN TVSGLVNAS LGLSMPANLS 1920
60 GGGNVGATVA GFVNNTQILN LGFGNVGSGN VGHGNIGDSN VGLGNLGNAN VGHGNIGSFN 1980
VFSGNRGSYN IGPANLGNYN IGLGNLGSYN FGFGNAGDFN LGFANS GSNN IGFANTGNNN 2040
IGIGLSGHNQ QGFGSWNSGT ANTGLFNSGT NNIGLFNSGT GNIGIGNSGI GNTGIGNPGV 2100
GNTGLGNSGT GNWGLWNP GT GNMGVANVT YNTGGYNVGS TNTGIANVGI ANTGSYNTGS 2160
TNTGSFNDGD FNTGFYNTGD YNTGFYNTGD VNTGAFIGN FSNAGFWQSD HQGQWGAHYA 2220
65 ITVPQIPLN FSLNIPVNIP IHLDFGTLAV NGFQIPAITL RALGVTHFSV GPIIVPRIAG 2280
TLPVIDINIG DPGGSSSIPI TITSGAGPVV IPLLDIPPAP GFGNSTTGPS SGFFNSGTGS 2340
SSGFGNVGAN NSGFWNTAFA GIGNSGLQNF GSLQSGWANL GNTVSGFYNT SAADFATPAN 2400

| | | | | | | | |
|----|-------------|-------------|------------|-------------|-------------|-------------|------|
| | LSGLSNVGAD | LTGVLRGPNG | STFNAGLANL | GQFNVGSANL | GSANLGSANL | GSANLGNSNV | 2460 |
| | GFGNIGNANI | GGANIGDFNV | GIANTGPGLT | AAVNNIGIGN | TGNYNIGVGN | TGNYNIGFGN | 2520 |
| | TGNNNIGIGL | SGDNQIGFGP | LNAGIANMGL | FNLGDNNFGM | ANAGNFNQGI | ANTGNNNIGL | 2580 |
| | FNTGNNNVGI | WLTGDGLSGF | SSLNSGAGNT | GFFNSGTANT | GLFNSGTGNT | GLFNSGTGNV | 2640 |
| 5 | GIGNMGTGGF | GVGLSGDSQV | GIGGTNSGSF | NIGLFNSGTG | NVGIGNSGTG | NVGIGNTG TG | 2700 |
| | NTGIGNSGNY | NTGLLNAGLV | NTGIANPGNH | NTGLFNIGTF | NTGIANPGHY | NTGSYNTGSY | 2760 |
| | NTGMANAGDY | GTGAFITGSM | NNGLLWRADR | QGLLAANYTI | TIERPAAFLN | VDIPVNIPIT | 2820 |
| | GDITNVSIPA | ITFPRIDASC | SVDIGILSGT | VLAPVGPITL | HGGDASAPLD | TPIEIDFGPS | 2880 |
| | PAINLNIGKP | DGSTVINIVG | GAGAGPISIP | IIDL RPAPGF | FNATTGPSSG | FLNWGAGSAS | 2940 |
| 10 | GLLNFGNNSG | LYNFATSSMG | NSGFQNYGSL | QSGWANLGNS | ISGIYNTGLG | APANVSGLLN | 3000 |
| | IGTNLAGWLQ | NGPTETTFVS | GLANLGFWNL | GSANIGNYNL | GSANIGVYNL | GSANIGDFNL | 3060 |
| | GSANIGDFNL | GSANIGSSNI | GFCNVGPGLT | AAIGNIGFGN | TGNGNIGIGN | TGTGNIGFGN | 3120 |
| | TGNGNIGIGL | TGDTMTGFGG | WNSGTGNIGL | FNSGTGNIGF | GNSGTGNWGI | GNSGDYNTGI | 3180 |
| | GNTGSTNSGF | FNTGLVNTGI | GNSGDYNTGL | FNAGNTNTGS | FNP GDYNTGG | FNP GNYNTGY | 3240 |
| 15 | FNP GNSNTGI | ANS GDVNTGA | FNSGNYSNGF | FWRGDYQGLG | GFAYQSAVSE | IPWSYDRFQH | 3300 |

<212> Type : PRT

<211> Length : 3300

SequenceName : SEQ ID 133

20 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv

25 <400> PreSequenceString :

| | | | | | | | |
|----|-------------|-------------|-------------|-------------|------------|-------------|------|
| | MNLVSTTSGM | SGFLNVGALG | SGVANVGNTI | SGIYNVGTSD | LSTPAVNSGL | ANIGTNIAGL | 60 |
| | LRDGAGTAAI | NLGLANHGNI | NVG FASLGGF | NFGGATIGHN | NVGIGNTGIF | DVGLANLGSY | 120 |
| | NIGFGNLGDD | NLGFNGFGSY | NIGFGNVGND | NLGFANAGGG | NIGFANTGSN | NVGFGNTGSN | 180 |
| | NVGIGLTGNG | QIGFGSFNSG | SGNIGLFNSG | SNNIGFFNSG | SGNFGIANS | SFNTGIGNTG | 240 |
| 30 | NTNTGLEFNSG | DVNTGAFNPG | SFNTGSFNTG | SFNTGGFNP | NTNTGYLNIG | NYNTGIAN TG | 300 |
| | DVDTGAFITG | NYSNGLFLSG | DYQGLVGLNL | VIDMPLPISL | GVNIPIDIPI | TASAGNITLM | 360 |
| | GVTIPPTGDI | VLSSIAGQRA | HFGPITIPNI | TVVGPTTTVA | IGGPNTAITI | TGGGAIRIPL | 420 |
| | ISIPAAPGFG | NSTTNPSSGF | FNTGAGGASG | FGNFGGANS | FWNLASATSG | ASGLLNVGAL | 480 |
| | GSGLANVGTT | VSGFYNTSTS | DLATPAFNSG | LANISTSIAG | LLRDSTGTMV | LNLGLANHGT | 540 |
| 35 | LNVGIANLGD | YNIGFANLGS | ANFGSANIGG | NNIGGANTGI | FDIGLANLGS | YNIGFGNFGD | 600 |
| | DNLGFGNLGS | YNVGFGNLGN | DNLGFANTGS | NNIGFANTGS | NNIGIGLTGD | GQIGFGSLNS | 660 |
| | GSGNIGLFNS | GSGNIGFFNS | GNGNVGIGNT | GTANFGLGNT | GSTNTGFFNS | GDVNTGIGNT | 720 |
| | GSFNTGSFNP | GDSNTGDFNP | GSYNTGLGNT | GDVDTGAFIS | GSYSNGFLWS | GNYQGLIGLH | 780 |
| | AALAIPEIAL | TFGVDIPIHI | PINIDAGVVT | LQGF SIVAAE | NNIDFTPIII | PTINITLPTA | 840 |
| 40 | AITVGGPTTS | IGITASAGIG | SITIPIIDIP | ATSGFGNSTT | SPSSGFFNSG | AGSASGFLNV | 900 |
| | VAGASGISGY | LNVGALGSGV | TNVGHTVSGF | YNASALDLVT | PAFASGLMRD | GMGTMTLNLG | 960 |
| | LANLGSNNAG | FGNTGIFDVG | VANLGNYNIG | FGNFGDDNLG | FANLGSYNIG | VANTGSNNIG | 1020 |
| | FANTGSNNIG | IGLTGTGQIG | IGALNSGSGN | IGLFNSGDGN | IGFFNSGTGN | FGIGNTG TGN | 1080 |
| | FGIGNSGSTS | TGLFNSGDGN | TGGFNPGNFN | TGNFNTGSFN | TGGFNAGNTN | TGHFNTGNYN | 1140 |
| 45 | TGIANTGDVS | TGAFISGNYS | NGILWRGDYQ | GLIGYSYALT | IPEIPAHLDV | NIPIDIPITG | 1200 |
| | SFTDLVVDNF | TIP IIGFESF | AFS FHIHTEP | DIGPIIVPSF | VLSVPTFAIA | VGGPTTAINI | 1260 |
| | SATAGLGPI | IPIIDIPAAP | GIGNSTTSPS | SGFFNTGAGT | ASGFGNVGGN | TSGLWNLASA | 1320 |
| | ASGVSGLLNV | GALGSGVANV | GNTISGIYNT | SPLDLGTPAF | GSGLANIAGL | LQGGAGTTIL | 1380 |
| | DLAGLGNNLV | GLANLGGSNF | GIGNTGIFNV | GFANVGNHNI | GLANLGNYSV | GFANS GNYHI | 1440 |
| 50 | GIANTGSANI | GFANTGSGNI | GIGLTGTGQI | GFGSFNSGSH | NIGLFNSGDG | NVGFFNSGTG | 1500 |
| | NVGIGNTGTA | NFGIANS GSF | NTGLGNTGST | NTGLFNPGNV | NTGVGNTGSI | NTGSINTGSF | 1560 |
| | NTGSTNTGSF | NLGDHNTGSF | NSGDYNTGYF | NAGDYNTGVA | NTGNVNTGAF | ISGNYSNGFF | 1620 |
| | WRGDYQGLIG | LSTTITIPEI | PYRYDLSVPI | DIPITGTVVA | TTPNSFTIPG | FQIRVLLGPA | 1680 |
| | AVLVNEMIGP | ITIDVNQVIA | IDSPIQQTIS | MVGTGGFGPI | PIGISIGGTP | GFGNSTTGPS | 1740 |
| 55 | SGFFHTGAGH | VSGFGNFAG | NMSGSGNFAG | GNSGFFNAGG | LGNSGLLNFG | ALQSGLANLG | 1800 |
| | NTISGVYNTS | TLDLATPAFG | SGIANIGANL | AGLFLDNTGN | LTLNFGVANQ | GGLNAGIGNL | 1860 |
| | GSVNIGFVNT | GDSNLGIGNL | GDLNFGGVNI | GGNNIGIANT | GIFDIGLANL | GSYNIGLANL | 1920 |
| | GDDNLGFGNA | GSYNIGFANF | GSDNLGFANT | GSYNIGFANT | GNNNIGVGLT | GNGQIGIGSL | 1980 |
| | NSGSMNIGLF | NSGSGNIGFF | NSGTGNVGIF | NTGTGNFGLA | NSGGFNTGIG | NAGSTNTGVF | 2040 |
| 60 | NPGDLNTGSF | NPGSFNTGGF | NPGSGNTGYL | NTGDYNTGVA | NTGDVDTGAF | ITGSYSNGFL | 2100 |
| | VSGDYQGLIG | LPLLGIPTVP | GYFNLTGGPS | SGFFNSGAGS | VSGFVNSGAG | LSGYLNTGAL | 2160 |
| | GSGVANVGNT | ISGWLNASAL | DLATPGFLSG | IGNFGTNLAG | FFRG | | 2204 |

<212> Type : PRT

<211> Length : 2204

65 SequenceName : SEQ ID 134

SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv

<400> PreSequenceString :

```
5  MSFVLIAP EF VTAAAGDLTN LGSSISAANA SAASATTQVL AAGADEVSAR IAALFGGFGL      60
   EYQAISAQVA AYHQR FVQAL STGAGAYASA EAAAAEQIVL GVINAPTQAL LGRPLIGDGA      120
   NATTPGGAGG AGGLLFGNGG AGAAGAPGQA GPGGPGAGLW GNNGPPGGAGG SGGGTGGAGG      180
   AGGWLFVGG AGGVGGAGGG TGGAGGPGGL IWGGGGAGGV GGAGGGTGGA GGRAELLFGA      240
   GGAGGAGTDG GPGATGGTGG HGGVGGDGGW LAPGGAGGAG GQGGAGGAGS DGGALGGTGG      300
10  TGGTGGAGGA GGRGALLLGA GGQGLGGAG GQGGTGAGG DGVLGGVGGT GGKGGVGGVA      360
   GLGGAGGAAG QLFSAGGAAG AVGVGGTGGQ GGAGGAGAAG ADAPASTGLT GGTGFAGGAG      420
   GVGGQGGNAI AGGINGSGA GGTGGQGGAG GMGGSGADNA SGIGADGGAG GTGGNAGAGG      480
   AGGAAGTGGT GGVVGAAGKA GIGGTGGQGG AGGAGSAGTD ATATGATGGT GFSGGAGGAG      540
   GAGGNTGVGG TNGSGGQGGT GGAGGAGGAG GVGADNPTGI GGTGGTGKKG GAGGAGGQGG      600
15  SSGAGGTNGS GGAGGTGGQG GAGGAGGAGA DNPTGIGGAG GTGGTGGAAG AGGAGGAIGT      660
   GGTGGA VGSV GNAGIGGTGG TGGVGGAGGA GAAAAAGSSA TGGAGFAGGA GGEAGGAGNS      720
   GVGGTNGSGG AGGAGGKGGT GGAGGSGADN PTGAGFAGGA GGTGGAAGAG GAGGATGTGG      780
   TGGVVGATGS AGIGGAGGRG GDGGDGASGL GLGLSGFDGG QGGQGGAGGS AGAGGINGAG      840
   GAGNGGDDGG DGATGAAGLG DNGGVGGDGG AGGAAGNNGN AGVGLTAKAG DGAAGNNGN      900
20  GGAGGAGGAG DNNFNGGQGG AGGQGGQGGG GASTTSINA NGGAGNNGT GGKGGAGGAG      960
   TLGVGSGGTT GGDGGDAGSG GGGGFGGAAG KAGGGNNGR GGDGGDGASG LGLGLSGFDG     1020
   GQGGQGGAGG SAGAGGINGA GGAGNGGDDG GDGATGAAGL GDNGGVGGDG GAGGAAGNNG     1080
   NAGVGLTAKA GDGAAGNNG NGGAGGAGGA GDNNFNGGQG GAGGQGGQGG LGGASTTSIN     1140
   ANGGAGNNGG TGKGGAGGA GTLVGSGSGG TGGDGGDAGS GGGGFGGAA GKAGGGNNGG     1200
25  VGGDGGEGAS GLGLGLSGFD GGQGGQGGAG GSAGAGGING AGGAGGTGGA GGDGAPATLI     1260
   GGPDDGGDGGQ GGIGGDGGNA GFGAGVPGDG GDGAGNAGFA GVPDGGIGG TGGAGGAGGA     1320
   GADGDPSIDG GQGGAGGHGG QGKGGGLNST GLASAASGDG GNGGAGGAGG NGGDGDGFIG     1380
   GSGGTGGTGG DAGVGLANT GGTAGNAGIG GAGGRGDDGG AGDSGALSQD GNGFAGGQGG     1440
   QGGVGGNAGA GGINGAGGTG GTGGAGDDGQ NGTTGVASEG GAGGQGGDGG QGGIGGAGGN     1500
30  AGFGAGVPGD GGIGGTGGAG GAGGAGADGD PSIDGGQGGG GGHGGQGGKG GLNSTGLASA     1560
   ASGDGGNNGA GGAGNGGDDG DFIGGSGGT GGTGGDAGVG GLANTGGTAG NAGIGGAGGR     1620
   GGDGGAGDSG ALSQDGNFAG GGQGGQGGVG GNAGAGGING AGGTGGTGGA GGDGQNGTTG     1680
   VASEGGAGGQ GGDGGQGGIG GAGGNAGFGA GVPDGGIGG TGGAGGAGGA GADGDPSIDG     1740
   GQGGAGGHGG QGKGGGLNST GLASAASGDG GNGGAGGAGG NGGAGGLGGG GGTGGTNGNG     1800
35  GLGGGGNNGG AGGAGGTPTG SGTGGTGGDG GDAGAGNNGG SATGVNGNNG GGDGGNNGDG     1860
   GNGAPGGFGG GAGAGGLGGS GAGGTDGDD GNGSPGTDG S      1901
```

<212> Type : PRT

<211> Length : 1901

SequenceName : SEQ ID 135

40 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv

45 <400> PreSequenceString :

```
45  MSLVIVAPET VAAAALDVAR IGSSIGAANA AAAGSTTSVL AAGADEVSAA IATLFGSHAR      60
   EYQAISTQVA AFHDRFAQTL SAAVGSYVSA EATNAAPLAT LEHNVLNALN APTQALLGRP      120
   LIGDGAAGAP GTGQAGGAGG ILWNGGAGG SGAPGQVGGG GGAAGLFGTG GAGGAGGAGA      180
   AGGAGGSGGW LLGNGGVGGA GGQSLGGAT GGAGGNAGLF GVGGTGGPGG PGGPGGVGGT      240
50  GGAGGLGCTL YGAGGHGGAG GPGPIGGVGG HGGVGGAGL LGVGGHGGAG GHGAEGVAGA      300
   AGEDLSPHGT SGGVGGDAGD GGTGGRGWL AGAGGAGGAG GVGGTGGAGG AGFSRALIVA      360
   GDNGGDPGAG GAGGTGGAGS TIGAHGAAGA SPTSGNNGA GNGAHFSSG GKAGGNGGAG      420
   GAGGLVGNNG AGGAGNGAP GAPPSSGDPN GGGGAGGAG GKGGDGGAQA GDGAGGAGG      480
   KGGNGNGAT GATGLNGLGA GADGTDGKG GNGGAGGGG AGGQGGKALA ATHQDGSMA      540
55  GGAGNGGAG GMGGDGGNGA KGTFDNGGDG VGGNGNGGS RGIGGAGGIG GAGSTAGADG      600
   ARGATPTSGG NGGTGGNGAN ATVAGGAGGA GKGGGNGLV GNGGAGGKG DGMAGVAGSS      660
   PTTAGESGTS GQNGGAGGAG GAGGRGDDFG GDGCTGGAGG NGANGANATT PGAKGGDGGH      720
   GPGPAQGGNG GQGGPGGLAG NLFQNGIQG VGGSGGKGA GGLAGDGGNG ANGNFAFGDG      780
   NGGHGNGGN PGAGGQGGSG GAGSTPGAAG AHGFTPTSGG DGGDGGNNGN SQVVGNGGD      840
60  GGNGGNGGSA GTGGNGGRG DGAFFGMSAN ATNPGENGN GNPGGNGGAG GAGGAGLNGG      900
   NGGAGNGGL GFGGNGAAG ANGVAVGAPG QPGGAGGHGG AGNGGAGGN GQGVVSDGA      960
   GGAGGAGDG GAPGDGANGG NGQAGAFAG GGGGRGDDG NAGNAGAGGP GGTGSTAGKA     1020
   GPAGSILHDG GNGGHGGHGA ASGNGGPGG HGGNGNGGT GANGNGGIG GTGGAGSTGA     1080
   KVLGTNEGD GGDGGRGGNG GRGNGGQGL TGAGNGGTG GTPGNGNGG NGASGLVTS      1140
65  PGDGGGGGRG GDAGRGGDAG LGSSGPGGT PGDWGTGGT GTGGTGQGA NGGLTGGRG      1200
   TGGNGNGNT GGTGGAGGTG GTGHNGSQPG MGGNGAGGF GNGFAGVGG RGMGGSGGT      1260
   GGTGDAGPFG TGTGGTGHHG QGGGGGFSI LLGLGGLGGL GSPGSIATGT AGGAGGGGGF     1320
```


GGLGGGEFV 1329
<212> Type : PRT
<211> Length : 1329
SequenceName : SEQ ID 136
SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

| | | | | | | | |
|----|-------------|------------|------------|-------------|-------------|------------|------|
| 10 | MSYVIATPEM | MATAAFDLAR | IGSQVSAASA | VAAMPTTEVV | AAGADEVSAG | IAALFSAHAQ | 60 |
| | EYQALSAQAA | AFHDQFVHTL | TAAARWYTAT | EIANAAAMRV | VLGAVNAPTQ | TLLGRPLIGD | 120 |
| | GAHGTA PGQP | GGAGGLLFGN | GGNGAAGAVG | QVGGAGGAAG | LFGIGGAGGA | GGAGAPGGTG | 180 |
| | GTGGWLAGGG | GVGGMGGAGG | GAGGAGGNAG | LFGNGGAGGA | GGAGGGAGGA | GGNAGWFGHG | 240 |
| 15 | GAGGVGGVGA | AGANGATPGQ | DGAAGVAGSD | DCAGGDGLAG | SDGGDGGAGG | VGGNGGRGGW | 300 |
| | LLNGGGAGGV | GGVGGAGGAG | AAGGAGGAGA | TGNGPAGIS | AAGGDGGAGG | NGGAGGNNGV | 360 |
| | GGAGGAGGSA | GLLGYVGRAG | DGGAGGGGGL | GGAPGDGGAG | GNGGSWLAAG | DGGAGGHGGD | 420 |
| | PGLGGAGGAG | GASGGAGARA | GANGLAAGND | GPVSGGNGGK | GGNGAHAPVA | GGHGGNGGAG | 480 |
| | GNGGLVGDGG | AGGHGGDGAA | GAGYADMTAI | FLGSSGTPGE | DGGNGGAGGA | GGAGGAHAGD | 540 |
| 20 | GGAGGAGGNG | GAGGAGGNGA | HGFNAVLVSD | GGNGGDGGAG | GRGGDGGAGG | AGGDAPAGRA | 600 |
| | GSQGVGGDGG | AGGAGGAPGN | GGSGGRGDMA | FKDGDGGAGG | DGGDPGAGGK | GGAGGAGATE | 660 |
| | GVTGATGATV | HSGGNGGKGG | NGADATVAGA | NGGKGGAGGN | GGLVGDGGAG | GDGGSGAAGA | 720 |
| | NGANVGEDGA | DGTLSCQPGE | GSEANGGQGG | VGGGGAGGAG | GDGGAGSSAL | GSGGNGGRGD | 780 |
| | AGQAGGAGGA | GGAGGAGGSV | SGDGGPGGKG | GAGGAGGAGA | SGGGGGKGAS | GADSAEAVGG | 840 |
| 25 | AGGKGGDGGV | GGVGGDGGPG | GDGGAGGAAP | AGQVGSHEVG | GVGGDGGGLG | AGGNGGDGGH | 900 |
| | GSDGGDGGDG | GDPGAGGLGG | LGDSGNGTR | AASGVDASDH | GPSSGNGGN | GGNGAQASVA | 960 |
| | GGAGGNGGDG | GNAGRVGDGG | AGGNGGDGAA | GANGANS GAP | GSDALALGQP | GGNGGQGDAG | 1020 |
| | QAGGAGGAGG | AGGAGGSVSG | DGGAGGNNGA | GGNGGVGASG | GAGARGANGI | DSIGGTGGAG | 1080 |
| | GGGGDGGAGG | VGGHGGDGGV | GGAAPSGTVG | SHGTGGVGGD | GGLGGAGGVG | GAGGNGGIGI | 1140 |
| 30 | TVGGAGGAGG | NGGDPGAGGR | GGLGGDSGNG | TSAANGVDAS | KHGPLTGGDG | GVGGNGAKAA | 1200 |
| | AAGGDGGQGG | DGGNAGLFGD | GGAGGDGADG | TAAEALGGDG | GAGGAGGKGG | DAGDIGDGGD | 1260 |
| | GGKGGDGAHG | ALGGLTVAGG | NGGAGGAGGA | GGAGGAFLGD | GGNGGAGGQG | GAGRGGSPGG | 1320 |
| | GGGVGGHGGG | GGDAGMNGGG | GTGGQGGNGA | AGGAGWSPDS | DLKGFDFGFDG | GSGGAGGDGG | 1380 |
| | AGGAGGTQTG | DGGDGGAGGL | GGAGGVGGNG | VDGFDINETT | GRDGGDGGDG | GYGGWGGAGG | 1440 |
| 35 | NGGAGGSAPA | GEVGNRGVGG | DGGDGGSGGD | AGNGGLGGDG | FTYLADFDGE | PGGDGGDGGD | 1500 |
| | GGWGRPGGQG | GFGSTSGAHG | KAGFGAPGGD | GDDGGNGGHG | GDNGSFADA | GDGGPGGNGG | 1560 |
| | NGGLGGAGRD | GGAPGGDGGD | GGTGGSGGFG | APPPRSIGGG | DGGDGGRGGD | GGRGAGGLTS | 1620 |
| | GGVGSSGESG | GSGNGRGDPG | SGSGSGEGGE | GGPSISVNVT | | | 1660 |

<212> Type : PRT
<211> Length : 1660
SequenceName : SEQ ID 137
SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

| | | | | | | | |
|----|------------|------------|------------|------------|------------|------------|------|
| 45 | MSFVLVSPET | VAAVATDLKR | IGASLAHENA | SAAASTTAVV | SAAADEVSTA | VAALFSQHAQ | 60 |
| | GYQAAAAQVA | AFHSRFVQAL | TAGAGAYAFA | EAANASPLQS | AMGAVSASAQ | TLLSRPLIGN | 120 |
| 50 | GANATTPGGN | GGDGGWLFGS | GGNGAPGAAG | QSGGNGGSAG | LWNGGAGGA | GGSGGAAGGN | 180 |
| | GGNGGWLFGA | GGTGGIGGTG | APGAMGGTGG | NGNGALLIG | GGGLGGAGGM | GGTGGGTGGT | 240 |
| | GGNGGNGALL | IGAGGVGGAG | GIGGQGTGAG | GAAGAGGTGG | NGGAGGLFMN | GGDGGAGGQG | 300 |
| | GDGAAGDAAA | SAGGTGGKGG | QGGDGGTGGA | GGAGPVLFGH | GGAGGMGGQG | GTGGMGGAGG | 360 |
| | DGTTVIAAGT | GEGGTGGAA | GAGGAAGARG | ALTSGGLAGG | VGAGGTGGTG | GTGGNGADAA | 420 |
| 55 | AVVFGANGD | PGFAGGKGGN | GGIGGAAVTG | GVAGDGGTGG | KGGTGGAGGA | GNDAGSTGNP | 480 |
| | GGKGGDGGIG | GAGGAGGAAG | TGNGGHAGNT | GDGGDGGTGG | NGNGTGTVN | GADNTLNPD | 540 |
| | PGGAGEPGGA | GGAGGAGGAA | GGPGGTGGTG | GNGGNGNGG | NGGNGNGGN | GGNAGNNSTN | 600 |
| | APVGEGGAG | GDGGAGGAGG | AANGGTAGSQ | GTGGVGGDGG | AGGNGGGGKA | GTGNSGNFGV | 660 |
| | DGEAGFSGGA | GGNGGVGGAA | GANGGTGGS | GNGGDGGAGG | IGGAGNGGIP | GTGTEPAGGT | 720 |
| 60 | GAKGGDGGDG | GAGGAGGNAG | GAGGQGGNAG | QGGAGGAGGN | AVIPGDGVGK | APHGDAGGSG | 780 |
| | GDGGKGGQGG | SGGTGGSGAP | IGGGAGGTGG | SGGHAGKGGG | GGIGAQTITI | TVPNGGNAG | 840 |
| | DGGNGGNAGA | GGNGSGDFG | GNTTSGASGS | GGNGGNAGTA | GSGGAGGTGG | TGLSGGNGGN | 900 |
| | GGNGGNGGDG | GNGAHGTVGA | QFVPATSLPT | PNGGAGGNGG | TGSNGGAPGP | AGAPGPTTGG | 960 |
| | NAGSQGIGGD | GGNGGDGGKG | GDGADAVNVV | FMPTEPQAAT | GTAGSAGDPT | GGNGGPPTPG | 1020 |
| 65 | SPMVAPPPT | PITQVQQGGD | GGAGGTGSTN | ANDGTATGGK | GEGGVGSIL | GGPGGNGGTG | 1080 |
| | GNASATGTNG | VANAGNGGKG | GDGGQFGAGG | NGGAGGSVTD | GSAGSTAGNG | GNGGNATNGT | 1140 |
| | IAGQPAGGNG | SAGGKGGDGG | NIAAGATGTA | GNGGNGGNGN | DGAVNAGTGG | SGGNGGNAGG | 1200 |

GGANGGDGGA GGAGGAGGRG GKGIDGGFGG DGGNGGSNNG TGAGGNGGNG GTGGVGSVGA 1260
AGGDGGNGGT GGFAGFGGTA GNGGSGGTGG AGGDGGTGGD GGNGVIAGGG GTGGNGGASG 1320
AGGAGGTGGF AGNGNAGGNG GTGGASEDGD NGNAGSGATG GTGGNGGTGG DGGAAAGLGGV 1380
A 1381

5 <212> Type : PRT
<211> Length : 1381
SequenceName : SEQ ID 138
SequenceDescription :

10 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

15 MSFVIATPEM LTAAATDLAK IGSTITAANT AAAAVAKVLP ASADEVSVAV AALFGTHAQE 60
YQTVSAQVAT FHDRFVQTLS AAASSYVAEE AVNVEQSLLA AVNAPTQALF GRPLIGNGAD 120
GSPGTGQAGG PGGILYNGGG NGGSGAPGQR GGAGGAAGLI GNGGNGGAGG VGTGAGAGGH 180
GGAGGWLYGN GGAGGFGGAG AVGGNGGAGG TAGLFGVGGG GGAGGNGIAG VTGTSASTPG 240
GSGTAGGAGG IGGNGGAGGA GGVLMGNGGN GGAGGEGGPG GAGGAGASGA HATNLGADGQ 300
AGGNGGNGGA GGTGGVGGPG GGHLLGLGG SHGAGGAGGS GGDGGAPGDG GNGATGTWGH 360
20 NLGAGGTGGN GGNPGAGGAG GAGGASVGS AHGANGAPGT TSTSGGNGGD GGKGADAISS 420
GQTGANGGRG GDGGQVNGG AGGAGGRGGA GGLFGSEAP GRPGGAGGTG GAGGNGGTQA 480
GDGGTGGAGG AGGDGGSGBA GSIGFNASAP GAAGSPGGNG GNGGPGGAGG EGGAGGLALA 540
ASGQNGSQGA GGDGAGGNG GTPGNGGHGA AGALGVNNGV GGAGGHGGDP GVGAGGQGG 600
SGSTPGANGA PGNTPTSGGN GNGGGRGADA TGFQGTGASG GRGGDGLVG NGGAGGAGGN 660
25 GSKGLPGLGR LGNPGLDGGT GNGGAGGSG GAWAGNGGTG GAGGTGGVGG TGGSGSDGVN 720
GSSAGADGHP GGTGGVGGTG GKGGDGGDGG AAPNGVAGSQ GPGGAGDGG TGGVGGNGGR 780
GIDGADGATA GARGQDGGAG GAGGKGRGG TGGPGGAGPA GTTGSQAGG NGGSGGTGGD 840
PGDGGNGANG SVFTNNGIGG NCGNCGNAGP SGAGGSGGAG STFGATGSSS SIHVNGGNGG 900
NCGNGDHLS GNGAAGGNGG NCGNGLRGS GGAGGHGGNG GNASRGMGGD GGTGGAGGNA 960
30 GQIGNGGAGG NGGDGGTGS DGNPAGITGSG GRGGDGGVGG QGGSVAGDGA DGGRGGAGGT 1020
GGTGLRGTG ATGATGTFDA GADGHGNGG TGGVGGTGA GGGGNGGAG GKALSPTGNN 1080
GSQAGGDGG AGGAGGTGGT GGDGGRGAHG TLFSSLAGTG GTGGNGGTGG TGGTGGAGGA 1140
GGTGSTLGAT GATGAAGRAG NGGVGSGGL GSAFGPGGTG GMGGAGGTST VSAGGDGGRG 1200
GFGGDGLDAS SGGNGGDGGH GGDGERTAGA GGRGGDGGKG ADPGGLFPIP GAGGKGGTGG 1260
35 TGGTAHLGPL AIIGSQGPQ QFGSPGADGR GGAGGAGGGG GAGGSF 1306
<212> Type : PRT
<211> Length : 1306
SequenceName : SEQ ID 139
SequenceDescription :

40 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

45 MSAAVAWDQ LAMELASAAA SFNSVTSGLV GESWLGPSSA AMAAAVAPYL GWLAAAAAQA 60
QRSATQAAAL VAEFEAVRAA MVQPALVAAN RSDLVSLVFS NFFGQNAPI AAIEAAYEQM 120
WAIDVSVMSA YHAGASAVAS ALTPFTAPPQ NLTDLPALQA AAPAAVVTAA ITSSKGVLAN 180
LSLGLANSF GQMGAAANLGI LNLGSLNPGG NNFGLGNVGS NNVGLGNTGN GNIGFGNTGN 240
GNIGFGLTGD NQQGFGGWNS GTGNIGLFNS GTGNIGIGNT GTGNFGIGNS GTSYNTGIGN 300
50 TGQANTGFFN AGIANTGIGN TGNNTGGSFN LGSFNTGDFN TGSSNTGFFN PGNLNTGVGN 360
TGNVNTGGFN SGNYSNGFFW RGDYQGLIGF SGTLTIPAG LDLNGLGSVG PITIPSITIP 420
EIGLGINSSG ALVGPINVP ITVPAIGLGI NSTGALVGP NIPPITLNSI GLELSAFQVI 480
NVGSISIPAS PLAIGLFGVN PTVGSIGPGS ISIQGLTPEI PAIPFFPGF PPDYVTVSGQ 540
IGPITFLSGG YSLPAIPLGI DVGGGLGPFT VFPDGYSLPA IPLGIDVGGG LGPFTVFPDG 600
55 YSLPAIPLGI DVGGGLGPFT VFPDGYSLPA IPLGIDVGA IGPLTTPPIT IPSIPLGIDV 660
SGSLGPINIP IEIAGTPGFG NSTTTPSSGF FNSGTGTSF FGNVSGSGS FWNIAAGNLGN 720
SGFLNVGPLT SGILNFGNTV SGLYNTSTLG LATSAPHSGV GNTDSQLAGF MRNAAGGTLF 780
NFGFANDGTL NLGNANLGDY NVGSGNVGSY NFGSGNIGNG SFGFGNIGSN NFGFGNVGSN 840
NLGFANTGPG LTEALHNIGF GNIGGNNGYF ANIGNGNIGF GNTGTGNIGI GLTGDNQVGF 900
60 GALNSGSGNI GFFNSGNGNI GFFNSGNGNV GIGNSGNYNT GLGNVGNANT GLFNTGNVNT 960
GIGNAGSYNT GSYNAGDTNT GDLNPGNANT GYLNLGDLNT GWGNIGDLNT GALISGSYSN 1020
GILWRGDYQG LIGYSDTLGI PAIPLSVEVN GGIGPIVVPD ITIPGIPLSL NALGGVGPIV 1080
VPDITIPGIP LSLNALGGVG PIVVPDITIP GIPLSLNALG GVGPIVVPDI TIPGIPLSLN 1140
ALGGVGPIVV PDIPTIPGIP SLNALGGVG PIVVPDITIP IPTINIRIP VNITLNEPFI 1200
65 NVAGIFTGYI GPIPLSTFVL GVTLAGGTLE SGIQGFVNP FGLNIPLSGA TNAVITIPGFA 1260
INPFGLNVPL SGGTSPVTIP GFAINPFGLN VPLSGGTSPV TIPGFTIPGS PLNLTANGGL 1320
GPINIPINIT SAPGFGNSTT TPSSGFFNSG DGSASGFGNV GPGISGLWNQ VPNALQGGVS 1380

GIYNVQQLAS GVANLGNTVS GFNNTSTVGH LTAAFNSGVN NIGQMLLGFF SPGAGP

1436

<212> Type : PRT

<211> Length : 1436

5 SequenceName : SEQ ID 140

SequenceDescription :

Sequence

10 <213> OrganismName : Mycobacterium tuberculosis H37Rv

<400> PreSequenceString :

| | | | | | | | |
|---------------|-------------|-------------|-------------|------------|-------------|------------|------|
| MEFPVLPPEI | NSVLMYSGAG | SSPLLA | AAAAA | WDGLAEELGS | AAVSFGQVTS | GLTAGVWQGA | 60 |
| AAAAMAAAAA | PYAGWLGSVA | AAAEAVAGQA | RVVVGVFEAA | LAATVDPALV | AANRRLVAL | | 120 |
| AVSNLLGQNT | PAIAAAEAEY | ELMWAADVAA | MAGYHSGASA | AAAALPAFSP | PAQALGGGVG | | 180 |
| 15 AFLTALFASP | AKALSLNAGL | GNVGNYNVGL | GNVGVFNLGA | GNVGGQNLGF | GNAGGTNVGF | | 240 |
| GNLGNGNVGF | GNSGLGAGLA | GLGNIGLGNA | GSSNYGFANL | GVGNIGFGNT | GTNNVGVGLT | | 300 |
| GNHLTGIGGL | NSGTGNIGLF | NSGTGNVGF | NSGTGNFGVF | NSGNYNTGVG | NAGTASTGLF | | 360 |
| NAGNFNTGVV | NVGSYNTGSF | NAGDTNTGGF | NPGGVNTGWL | NTGNTNTGIA | NSGNVNTGAF | | 420 |
| ISGNFNNVGL | WVG DYQGLFG | VSAGSSIPAI | PIGLVLNGDI | GPITIQPIPI | LPTIPLSIHQ | | 480 |
| 20 TVNLGPLVVP | DIVIPAFGGG | IGIPINIGPL | TITPITLFAQ | QTFVNQLPFP | TFSLGKITIP | | 540 |
| QIQTFDSNGQ | LVSFIGPIVI | DTTIPGPTNP | QIDL TIRWDT | PPITLFPNGI | SAPDNPLGLL | | 600 |
| VSVSISNPGF | TIPGFSVPAQ | PLPLSIDIEG | QIDGFSTPPI | TIDRIPLTVG | GGVTIGPITI | | 660 |
| QGLHIPAAPG | VGNTTTAPSS | GFFNSGAGGV | SGFCNVGAGS | SGWWNQAPSA | LLGAGSGVGN | | 720 |
| VGTLGSGVLN | LSGGISGFYN | TSVLPFGTPA | AVSGIGNLGQ | QLSGVSAAGT | TLRSMLAGNL | | 780 |
| 25 GLANVGNFNT | GFGNVGDVNL | GAANIGGHNL | GLGNVGDGNL | GLGNIGHGNL | GFANLGLTAG | | 840 |
| AAGVGNVGF | NAGINNYGLA | NMGVGNIGFA | NTGTGNIGIG | LVGDHRTGIG | GLNSGIGNIG | | 900 |
| LFNSGTGNVG | FFNSGTGNFG | IGNSGRFNTG | IGNSGTASTG | LFNAGSFSTG | IANTGDYNTG | | 960 |
| SFNAGDTNTG | GFNPGGINTG | WFNTGHANTG | LANAGTFGTG | AFMTGDYSNG | LLWRGGYEGL | | 1020 |
| VGVRVGPTIS | QFPVTVHAIG | GVGPLHVAPV | PVPAVHVEIT | DATVGLGPFT | VPPISIPSLP | | 1080 |
| 30 IASITGSVDL | AANTISPIRA | LDPLAGSIGL | FLEPFRLSDP | FITIDAFQVV | AGVLFLENII | | 1140 |
| VPGLTVSGQI | LVTPTPIPLT | LNLDTTPWTL | FPNGFTIPAQ | TPVTVGMEVA | NDGFTFFPGG | | 1200 |
| LTFPRASAGV | TGLSVGLDAF | TLLPDGFTLD | TVPATFDGTI | LIGDIPPIII | DVPAVPGFGN | | 1260 |
| TTTAPSSGFF | NTGGGGGSGF | ANVGAGTSGW | WNQGHDLVAG | AGSGVANAGT | LSSGVLNVGS | | 1320 |
| GISGWYNTST | LGAGTPAVVS | GIGNLGQQLS | GFLANGTVLN | RSPIVNIGWA | DVGAFNTGLG | | 1380 |
| 35 NVGDLNWGAA | NIGAQNGLGL | NLGSGNVGFG | NIGAGNVGFA | NSGPAVGLAG | LGNVGLSNAG | | 1440 |
| SNNWGLANLG | VGNIGLANTG | TGNIGIGLVG | DYQTGIGGLN | SGSGNIGLFN | SGTGNVGFEN | | 1500 |
| TGTGNFGLFN | SGSFNTGIGN | SGTGSTGLFN | AGNFNTGIAN | PGSYNTGSFN | VGDTNTGGFN | | 1560 |
| PGDINTGWFN | TGIMNTGTRN | TGALMSGTDS | NGMLWRGDHE | GLFGLSYGIT | IPQFPPIRIT | | 1620 |
| TGGIGPIVIP | DTTILPPLHL | QITGDADYSF | TVPDIPIPAI | HIGINGVTV | GFTAPEATLL | | 1680 |
| 40 SALKNNGSFI | SFGPITLSNI | DIPPMDFTLG | LPVLGPITGQ | LGPIHLEPIV | VAGIGVPLEI | | 1740 |
| EPIPLDAISL | SESIPIRIPV | DIPASVIDGI | SMSEVVPIDA | SVDIPAVTIT | GTTISAIPLG | | 1800 |
| FDIRTSAGPL | NIPIIDIPAA | PGFGNSTQMP | SSGFFNTGAG | GGSGIGNLGA | GVSGLLNQAG | | 1860 |
| AGSLVGTLTG | LGNAGTLASG | VLNSGTAISG | LFNVSTLDAT | TPAVISGFSN | LGDHMSGVSI | | 1920 |
| DGLIAILTFP | PAESVFDQII | DAAIAELQHL | DIGNALALGN | VGGVNLGLAN | VGEFNLGAGN | | 1980 |
| 45 VGNINVGAGN | LGGSNLGLGN | VGTGNLGFGN | IGAGNFVFGN | AGLTAGAGGL | GNVGLGNAGS | | 2040 |
| GSWGLANVGV | GNIGLANTGT | GNIGIGLTGD | YRTGIGGLNS | GTGNLGLFNS | GTGNIGFFNT | | 2100 |
| GTGNFGLFNS | GSYSTGVGNA | GTASTGLFNA | GNFNTGLANA | GSYNTGSLNV | GSFNTGGVNP | | 2160 |
| GTVNTGWENT | GHTNTGLFNT | GNVNTGAFNS | GSFNNGALWT | GDYHGLVGFS | FSIDIAGSTL | | 2220 |
| LDLNETLNLG | PIHIEQIDIP | GMSLFDVHEI | VEIGPFPTIPQ | VDVPAIPLEI | HESIHMDPIV | | 2280 |
| 50 LVPATTIPAQ | TRTIPLDIPA | SPGSTMTLPL | ISMREFEGEDW | ILGSTAAIPN | FGDPFPAPTQ | | 2340 |
| GITIHTGPGP | GTTGELKISI | PGFEIPQIAT | TRFLLDVNIS | GGLPAFTLFA | GGLTIPTNAI | | 2400 |
| PLTIDASGAL | DPITIFPGGY | TIDPLPLHLA | LNLTVPDSSI | PIIDVPPTPG | FGNTTATPSS | | 2460 |
| GFFNSGAGGV | SGFGNVGSNL | SGWWNQAASA | LAGSGSGVLN | VGTLGSGVLN | VSGVSGIYN | | 2520 |
| TSVLPPLGTPA | VLSGLGNVGH | QLSGVSAAGT | ALNQIPILNI | GLADVGNFNV | GFGNVGDVNL | | 2580 |
| 55 GAANLGAQNL | GLGNVGTGNL | GFANVGHGNI | GFGNSGLTAG | AAGLGNTGFG | NAGSANYGFA | | 2640 |
| NQGVNRNIGLA | NTGTGNIGIG | LVGDNL TGIG | GLNSGAGNIG | LFNSGTGNIG | FFNSGTGNFG | | 2700 |
| IGNSGSFNTG | IGNSGTGSTG | LFNAGSFNTG | VANAGSYNTG | SFNAGDTNTG | GFNPGTINTG | | 2760 |
| WFNTGHTNTG | IANSGNVGTG | AFMSGNFNSG | LLWRGDHEGL | FSLFYSLDVP | RITIVDAHLD | | 2820 |
| GGFGPVVLP | IPVPAVNAHL | TGNVAMGAF | IPQIDIPALT | PNITGSAAFR | IVVGSVRIPP | | 2880 |
| 60 VSVIVEQIIN | ASVGAEMRID | PFEMWTQGTN | GLGITFYSGF | SADGSPYATG | PLVFGAGTSD | | 2940 |
| GSHLTISASS | GAFTTPQLET | GPITLGFQVP | GSVNAITLFP | GGLTFPATSL | LNL DVTAGAG | | 3000 |
| GVDIPAITWP | EIAASADGSV | YVLASSIPLI | NIPPTPGIGN | STITPSSGFF | NAGAGGSGGF | | 3060 |
| GNFGAGTSGW | WNQAHTALAG | AGSGFANVGT | LHSGVLNLGS | GVSGIYNTST | LGVGTPALVS | | 3120 |
| GLGNVGHQLS | GLLSGGS AVN | PVTVLNIGLA | NVGSNAGFG | NVGEVNLGAA | NLGAHNLGFG | | 3180 |
| 65 NIGAGNLGFG | NIGHGNVGVG | NSGLTAGVPG | LGNVGLGNAG | GNNWGLANVG | VGNIGLANTG | | 3240 |
| TGNIGIGLTG | DYQTGIGGLN | SGAGNLGLFN | SGAGNVGFEN | TGTGNFGLFN | SGSFNTGVGN | | 3300 |
| SGTGSTGLFN | AGSFNTGVAN | AGSYNTGSFN | VGDTNTGGFN | PGSINTGWLN | AGNANTGVAN | | 3360 |

5 AGNVNTGAFV TGNFSNGILW RGDYQGLAGF AVGYTLPLFP AVGADVSGGI GPITVLPPIH 3420
IPPIPVGFAA VGGIGPIAIP DISVPSIHLG LDPVHVHSGI TVNPITVRTP PVLVSYSQGA 3480
VTSTSGPTSE IWVKPSFFPG IRIAPSSGGG ATSTQGAYFV GPISIPSGTV TFPGFPTIPLD 3540
PIDIGLPVSL TIPGFTIPGG TLIPTLPLGL ALSNGIPPVD IPAIVLDRIL LDLHADTTIG 3600
PINVPPIAGFG GAPGFGNSTT LPSSGFFNTG AGGGSGFSNT GAGMSGLLNA MSDPLLGSAS 3660
GFANFGTQLS GILNRGAGIS GVYNTGALGV VTAADVSGFG NVGQQLSGLL FTGVGP 3716

<212> Type : PRT

<211> Length : 3716

10 SequenceName : SEQ ID 141
SequenceDescription :

Sequence

15 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

MNFPVLPPEI NSVLMYSGAG SSPLLA AAAA WDGLAEELGS AAVSFGQVTS GLTAGVWQGA 60
AAAAMAAAAA PYAGWLGSVA AQAVAVAGQA RAAVAAFEAA LAATVDPAAV AVNRMAMRAL 120
AMSNLLGQNA AAIAAVEAEY ELMWAADVAA MAGYHSGASA AAAALPAFSP PAQALGGGVG 180
20 AFLNALFAGP AKMLRLNAGL GNVGNYNVGL GNVGIFNLGA ANVGAQNLGA ANAGSGNFGE 240
GNIGNANFGF GNSGLGLPPG MGNI GLNAG SSNYGLANLG VGNIGFANTG SNNIGIGLTG 300
DNLTGIGGLN SGTGNLGLFN SGTGNIGFFN SGTGNFGVFN SGSYNTGVGN AGTASTGLFN 360
VGGFNTGVAN VGSYNTGSFN AGTNTTGGFN PGNVNTGWLN TGNTNTGIAN SGNVNTGAFI 420
SGNFSNGVLW RGDYEGLWGL SGGSTIPAIP IGLELNGGVG PITVLPPIQIL PTIPLNIHQ 480
25 FSLGPLVVPD IVIPAFGGGT AIPISVGPIIT ISPITLFPAQ NFNTTFFVGP FFGLGVVNIS 540
GIEIKDLAGN VTLQLGNLNI DTRINQSFPV TVNWSTPAVT IFPNGISIPN NPLALLASAS 600
IGTLGFTIPG FTIPAAPLPL TIDIDGQIDG FSTPPITIDR IPLNLGASVT VGPILINGVN 660
IPATPGFGNT TTAPSSGFFN SGDGGVSGFG NFGAGSSGWW NQAQTEVAGA GSGFANFGSL 720
GSGVLNFGSG VSGLYNTGGL PPGTPAVVSG IGNVGEQLSG LSSAGTALNQ SLIINLGLAD 780
30 VGSVNVGFGN VGDFNLGAAN IGDLNVGLGN VGGGNVGFN IGDFNGLGN AGLAAGLAGV 840
GNIGLGNAGS GNVGFGNMGV GNIGFGNTGT NNLGIGLTGD NQTGIGGLNS GAGNIGLFNS 900
GTGNVGLFNS GTGNFGLFNS GSFNTGIGNG GTGSTGLFNA GNFTGVANP GSYNTGSFNV 960
GDTNTGGFNP GSINTGWENT GNANTGVANS GNVDTGALMS GNFSNGILWR GNFEGLFGLN 1020
VGITIEFPPI HWTSTGGIGP IIPDITILP PIHLGLTGQA NYGFAVPDIP IPAIHIDFDG 1080
35 AADAGFTAPA TTLALSALGIT GQFRFGPIITV SNVQLNPFNV NLKLQFLHDA FPNEFPDPTI 1140
SVQIQVAIPL TSATLGLLAL PLQQTIDAIE LPAISFSQSI PIDIPPIDIP ASTINGISMS 1200
EVVPIDVSVD IPAVTITGTR IDPIPLNFDV LSSAGPINIS IIDIPALPGF GNSTELPSSG 1260
FFNTGGGGGS GIANFGAGVS GLLNQASSPM VGTLSGLGNA GSLASGLVNS GVDISGMFNV 1320
STLGSAPAVI SGFGNLGNHV SGVSDGLLA MLTSGGSGGS GQPSIIDAAI AELRHLNPLN 1380
40 IVNLGNVGSY NLGFANVG DV NLGAGNLGNL NLGGGNLGGQ NLGLGNLGDG NVGFGNLGHG 1440
NVGFGNSGLG ALPGIGNIGL GNAGSNNVGF GNMGLGNIGF GNTGTNNLGI GLTGDNQTGF 1500
GGLNSGAGNL GLFNSGTGNI GFFNTGTGNW GLFNSGSYNT GIGNSGTGST GLFNAGSFNT 1560
GLANAGSYNT GSLNAGNTNT GGFNPGNVNT GWFNAGHTNT GGFNTGNVNT GAFNSGSFNN 1620
GALWTGDHGG LVGFSYSIEI TGSTLVDINE TLNLGPVHID QIDIPGMSLF DIHELNVNIGP 1680
45 FRIEPIDVPA VVLDIHETMV IPPIVFLPSM TIGGQTYTIP LDTPPAPAPP PFRPLPLFVN 1740
ALGDNWIVGA SNSTGMSGGF VTAPTQGILI HTGPSSATTG SLALTLPVT IPTITTSPIP 1800
LKIDVSGGLP AFTLFPGLLN IPQNAIPLTI DASGVLDPIIT IFPGGFTIDP LPLSLALNIS 1860
VPDSSVPIII VPPTPGFGNA TATPSSGFFN SGAGGVSGFG NFGAGSSGWW NQAHAALAGA 1920
GSGVLNVGTL NSGVNLVGS G ISGLYNTAIV GLGTPALVSG AGNVGQQLSG VLAAGTALTQ 1980
50 SPIINLGLAD VGNYNLGLGN VGDFNLGAAN LGDLNLGLGN IGNANVGFGN IGHGNVGFN 2040
SGLGAALGIG NIGLGNAGST NVGLANMGV NIGFANTGTN NLGIGLTGDN QTGIGGLNSG 2100
AGNIGLFNSG TGNIGFFNSG TGNWGLFNSG SFNTGIGNSG TGSTGLFNAG GFTTGLANAG 2160
SYNTGSFNVG DTNTGGFNPG SINTGWENTG NANTGIANSG NVDTGALMSG NFSNGILWRG 2220
NYEGLFSYSY SLDVPRITIL DAHFTGAFGP VVPPPIPVLA INAHLTGNA MGAFTIPQID 2280
55 IPALNPNTVG SVGFGPIAVP SVTIPALTA RAVLDMAASV GATSEIEPFI VWTSSGAIGP 2340
TWYSVGRIYN AGDLFVGNI ISGIPTLSTT GPVHAVFNAA SQAFNTPALN IHQIPLGFQV 2400
PGSIDAITLF PGGLTFPANS LLNLDVFGVT PGATIPAITF PEIPANADGE LYVIAGDIPL 2460
INIPPTPGIG NTTTTVPSSGF FNTGAGGSG FGNFGANMSG WVNQAHTALA GAGSGIANVG 2520
TLHSGVLNLG SGLSGIYNTS TLPLGTPALV SGLGNVGDHL SGLLASNVGQ NPITIVNIGL 2580
60 ANVGNGNVGL GNIGNLNLGA ANIGDVNLGF GNIGDVNLGF GNIGGNGVGF GNIGDANFGF 2640
GNSGLAAGLA GMGNIGLGA GSGNVGWANM GLGNIGFGNT GTNNLGIGLT GDNQSGIGGL 2700
NSGTGNIGLF NSGTGNIGFF NSGTANFGLF NSGSYNTGIG NSGVASTGLV NAGGFNTGVA 2760
NAGSYNTGSF NAGDTNTGGF NPGSTNTGWF NTGNANTGVA NAGNVNTGAL ITGNFSNGIL 2820
WRGNYEGLAG FSFGYPIPLF PAVGADVTDG IGPATIIPI HIPSIPLGFA AIGHIGPISI 2880
65 PNIAIPSIHL GIDPTFDVGP ITVDPIITLI PGLSLDAVS EIRMTSGSSS GFKVRPSFSF 2940
FAVGPDGMPG GEVSILQPFV VAPINLNPTT LHFPGFPTIP GPIHIGLPLS LTIPGFTIPG 3000
GTLIPQLPLG LGLSGGTPPF DLPTVVIDRI PVELHASTTI GPVSLPIFGF GGAPGFGNDT 3060

TAPSSGFFNT GGGGSGGFSN SGSGMSGVLN AISDPLLGS A SGFANFGTQL SGILNRGAGI 3120
SGVYNTGTLG LVTSAFVSGF MNVGQQLSGL LFAGTGP 3157

<212> Type : PRT

<211> Length : 3157

5 SequenceName : SEQ ID 142
 SequenceDescription :

Sequence

10 <213> OrganismName : Mycobacterium tuberculosis H37Rv
 <400> PreSequenceString :
MSFVVMPPPEI NSLLIYTGAG PGPLLAASAAA WDELAELGSA AAAAFGSVTS GLVGGIWQGP 60
SSVAMAAAAA PYAGWLSAAA ASAESAAGQA RAVVGVFEAA LAETVDPFVI AANRSRLVSL 120
ALSNLFGQNT PAIAAAEFDY ELMWAQDVAA MLCYHTCASA AAEALAPFGS PLASLAAAAE 180
15 PAKSLAVNLG LANVGLFNAG SGNVGSYNVG AGNVGSYNVG GGNIGGNNVG LGNVGWGNFG 240
 LGNSGLTPGL MGLGNIGFGN AGSYNFGLAN MGVGNIGFAN TGSGNFGIGL TGDNLTGFGG 300
 FNTGSGNVGL FNSGTGNVGF FNSGTGNWGV FNSGSYNTGI GNSGIASSTGL FNAGGFNTGV 360
 VNAGSYNTGS FNAGEANTGG FNPGSVNTGW LNTGDINTGV ANSGDVNTGA FISGNYSNGV 420
 LWRGDYQGLL GFSSGANVLP VIPLSLDING GVGAITIEPI HILPDIPINI NETLYLGPLV 480
20 VPPINVPAIS LGVGIPNISI GPIKINPITL WPAQNFNQTI TLAWPVSSIT IPQIQQVALS 540
 PPSPIPTTLIG PIHINTGFSI PVTFSYSTPA LTLFPVGLSI PTGGPLTLTL GVTAGTEAFT 600
 IPGFSIPEQP LPLAINVIGH INALSTPAIT IDNIPLNLHA IGGVGPVDIV GGNVPASPGF 660
 GNSTTAPSSG FFNTGAGGVS GFGNVGAHTS GWFNQSTQAM QVLPQTVSGY FNSGTLMSGI 720
 GNVGTQLSGM LSGGALGGNN FGLGNIGFDN VGFGNAGSSN FGLANMGIGN IGLANTGNGN 780
25 IGIGLSGDNL TGFGGFNSSG ENVGLFNSTG GNVGFFNSTG GNLGVFNSSG HNTGFFLTGN 840
 NINVLAPFTP GTLFTISEIP IDLQVIGGIG PIHVQPIDIP AFDIQTGGF IGIREFTLPE 900
 ITIPAIPHV TGTVGLEGFH VNPAFVLFQO TAMAETADP VVLPDPFITI DHYGPPLGPP 960
 GAKFPSSGSFY LSISDLQING PIIGSYGGPG TIPGPFGATF NLSTSSLALF PAGLTVPDQT 1020
 PVTVNLTGGL DSITLFPGGL AFFENPVVSL TNFSVGTGGF TVFPQGFTVD RIPVDLHTTL 1080
30 SIGPFPFRWD YIPPTPANGP IPAVPGGFGF TSGLFPFHFT LGGGIGPISI PTTTVVDALN 1140
 PLLTVTGNLE VGPFTVPDIP IPAINFGLDG NVNVSFNAPA TTLLSGLGIT GSIDISGIQI 1200
 TNIQTQPAQL FMSVGQTLFL FDFRDGIELN PIVIPGSSIP ITMAGLSIPL PTVSESIPLN 1260
 FSFGSPASTV KSMILHEILP IDVSINLEDA VFIPATVLPV IPLNVDTVITP VGPINIPIT 1320
 EPGSGNSTTT TSDPFSGLAV PGLGVGLLGL FDGSIANNLI SGFNSAVGIV GPNVGLSNLG 1380
35 GGNVGLGNVG DFNLAGNVG GFNVGGGNIG GNVVGLGNVG FGNVGLANSV LTPGLMGLGN 1440
 IGFGNAGSYN FGLANMGVGN IGFAANTGSGN FGIGLTGDNL TGFGGFNSTG GNVGLFNSTG 1500
 GNVGFFNSTG GNWGVFNSSG YNTGIGNSGI ASTGLFNAGG FNTGVVNAGS YNTGSFNAGQ 1560
 ANTGGFNPGS VNTGWLNTGD INTGVANSND VNTGAFISGN YSNGAFWRGD YQGLLGFSYR 1620
 PAVLPQTPFL DLTLTGGLGS VVIPAIDIPA IRPEFSANVA IDSFTVPSIP IPQIDLAATT 1680
40 VSVGLGPITV PHLDIPRPV TLNLYFGSQP GGPKKIGPIT GLFNTPIGLT PLALSQIVIG 1740
 ASSSQGTITA FLANLPFSTP VVTIDEIPLL ASITGHSEPV DIFPGGLTIP AMNPLSINLS 1800
 GGTGAVTIPA ITIGEIPFDL VAHSTLGPVH ILIDLPAVPG FGNTTGAPSS GFFNSGAGGV 1860
 SGFGNVGAMV SGGWNQAPSA LLGGGSGVFN AGTLHSGVLN FGSGMSGLEN TSVLGLGAPA 1920
 LVSGLSVGQ QLSGLLASGT ALHQGLVLNF GLADVGLGNV GLGNVGDFNL GAGNVGGFNV 1980
45 GGGNIGGNNV GLGNVGWGNF GLGNSGLTPG LMGLGNIGFG NAGSYNFGLA NMGVGNIGFA 2040
 NTGSGNFGIG LTGDNLTGFG GFNTGSGNVG LFNSGTGNVG FFNSGTGNWG VFNSGSYNTG 2100
 IGNSGIASSTG LFNAGGFNTG VVNAGSYNTG SFNAGQANTG GFNPGSVNTG WLNTGDINTG 2160
 VANSQDVNTG AFISGNYSNG AFWRGDYQGL LGFSYTSTII PEFTVANIHA SGGAGPIIVP 2220
 SIQFPAIPLD LSATGHIGGF TIPPVSISPI TVRIDPVFDL GPITVQDITI PALGLDPATG 2280
50 VTVGPIFSSG SIIDPFSITL LGFINVNVA IQTAPSEILP FTVLLSSLGV THLTPEITIP 2340
 GFHIPVDPIH VELPLSVTIG PFVSPEITIP QLPLGLALSG ATPAFAPPLE ITIDRIPVVL 2400
 DVNALLGPIV AGLVIPPVPG FGNTTAVPSS GFFNIGGGGG LSGFHNLGAG MSGVLNAISD 2460
 PLLGSASGFA NFGTQLSGIL NRGADISGVY NTGALGLITS ALVSGFGNVG QQLAGLIYTG 2520
TGP 2523

55 <212> Type : PRT

<211> Length : 2523

 SequenceName : SEQ ID 143
 SequenceDescription :

Sequence

60 <213> OrganismName : Mycobacterium tuberculosis H37Rv
 <400> PreSequenceString :
MSFVIAVPEA LTMAASDLAN IGSTINAANA AAALPTTGVV AAAADEVSAA VAALFGSYAQ 60
65 SYQAFGAQLS AFHAQFVQSL TNGARSYVVA EATSAAPLQD LLGVVNAPAQ ALLGRPLIGN 120
 GANGADGTGA PGGPGGLLLG NGGNGGSGAP GQPGGAGGDA GLIGNGGTGG KGGDGLVSGG 180
 AAGGVGGRGG WLLGNNGGTGG AGGAAGATLV GGTGGVGGAT GLIGSGGFGG AGGAAAGVGT 240

5 TGGVGGSSGV GGVFGNGGFG GAGGLGAAGG VGGAASYFGT GGGGGVGGDG APGGDGGAGP 300
LLIGNGGVGG LGGAGAAGGN GGAGGMLLGD GGAGGQGGPA VAGVLGGMPG AGGNNGNANW 360
FGSGGAGGQG GTGLAGTNGV NPGSIANPNT GANGTDNSGN GNQTGGNGGP GPAGGVGEAG 420
GVGGQGGGLGE SLDGNDGTGG KGGAGGTAGT DGGAGGAGGA GGIGETDGS A GGVATGGEGG 480
DGATGGVDGG VGGAGGKGGQ GHNTGVGDAF GGDGGIGGDG NGALGAAGGN GGTGGAGGNG 540
GRGGMLIGNG GAGGAGGTGG TGGGGAAGFA GGVGGAGGEG LTDGAGTAEG GTGGLGGLGG 600
VGGTGGMGGG GGVGGNGGAA GSLIGLGGGG GAGGVGGTGG IGGIGGAGGN GGAGGAGTTT 660
GGGATIGGGG GTGGVGGAGG TGGTGGAGGT TGGSGGAGGL IGWAGAAGGT GAGGTGGQGG 720
LGGQGGNGGN GGTGATGGQG GDFALGGNGG AGGAGGSPGG SSGIQGNMGP PGTQGADG 778

10 <212> Type : PRT
<211> Length : 778
SequenceName : SEQ ID 144
SequenceDescription :

15 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

20 PQGADGNAGN GGDGGVGGNG GNGADNTTAA AAGTTGGAGG AGGAGGTGGT GGAAGTGTGG 60
QQNGGNGGN GGTGGKGGTG GDGALAGSSG GAGGKGGNGG DAKKAGTGA PGTAGTGGDG 120
GKGGNGGIGA AGTTGPVGTG ASGGTGGSGG AGGTGGDGA ANGGTAGAGG AGNGGKGGD 180
GGAGVTSSTA GNSGGAGGSG GKGGDAGAGG AGATPGANGI AGNGGDGGDG AAGAVGISGA 240
TGAGDGGHGG TGAAGGNGGT GGAGGSGIDG VGGGTGGTGG NGNGAIGGA GGDAGGSGNS 300
25 GNGGGIGGKG GNAGAGGAAG SNGGTVGANG TGGDGGNGGA AGAATAGSNG GAGTGSAGGN 360
GGTGGRGGSG GAGGDGIGGV GGGKGGNGAD GEVGGAGGAG GSGPNTSPGG NGGQGGQGGG 420
GGAGGAAGAG GAGGGANGTA GNGGQGGAGG TGGAGAASSA TNGSGGAGG TGGDGGSGGA 480
GGTGGAGGTG GAAGDGGQGG QGGAGGGAGG QGGAGGAGGT GGNGGNITGG TAGTAGAAGN 540
GGAAGKGGAG GQGGTGGGTG QGGAGGDDG AGGTGGDRTV GGGTVPAGSG GQGGNAGGGG 600
30 AGGQGGADGG SGGDGGDAGT GGNGGNGGNR NSGNGTGGAG GNGGGGANGG AGGAGGSGGG 660
TGGNGGAGGD AGDAGNGGNG NGTGNNGNGG NGGIAGMGN GGAGTGSNG GNGSGGNGG 720
NAGMGNSGT GSGDGGAGGN GGAAGTGGTG GDGGLTGTGG TGGSGGTGGD GGNGGNGADN 780
TANMTAQAGG DGGNGGDGGF GGGAGAGGGG LTAGANGTGG QGGAGGDGNG GAIGHGPLT 840
DDPGGNGGTG GNGGTGGTGG AGIGSLGGGT GGDGGNGGNG GTGEGGEGVG GAGGTGGAAG 900
35 NGGDGGTGGT GGGDGGAGGT GGTGGTGGLG DPRVGGSGGD GGTGGSGGAA GNGGNGGNAG 960
AGGNGNGGTG GAGGIGGTGG NGGDAEPGVP PGAGGAGGAG TTGGKGGTGG NGS GTSGGT 1020
GGDGGTGGGG GNGGTGWNGG KGD TSGGGA GDGKAPAGG TGGAGGDGGA GGKGGSGGV 1079

40 <212> Type : PRT
<211> Length : 1079
SequenceName : SEQ ID 145
SequenceDescription :

45 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

50 MVMSLMVAPE LVAAAAADLT GIGQAISAN AAAAGPTTQV LAAAGDEVSA AIAALFGTHA 60
QEYQALSARV ATFHEQFVRS LTAAGSAYAT AEAANASPLQ ALEQQVLGAI NAPTQLWLGR 120
PLIGDGVHGA PGTGQPGAG GLLWNGGNG GSGAAGQVGG PGGAAGLFGN GSGSGSGGAG 180
AAGGVGGSGG WLNGNGGAGG AGGTGANGGA GGNLWLFAG GSGGAGTNGG VGGSGGFVYG 240
NGGAGGIGGI GGIGGNGDA GLFGNGGAGG AGAAGLPGAA GLNGGDGSDG GNGGTGGNGG 300
RGGLLVGNGG AGGAGGVGGD GKGAGGDPG FAVNNGAGGN GGHGNGPGVG GAGGAGLLA 360
GAHGAAGATP TSGGNGDGG IGATANSPLQ AGGAGGNGGH GGLVNGGTG GAGGAGHAGS 420
55 TGATGTALQP TGGNGTNGGA GGHGNGGNG GAQHGDGCVG GKGGAGGSGG AGGNGFDAAT 480
LGSPGADGGM GGNGGKGGDG GKAGDGGAGA AGDVT LAVNQ GAGGDGNGG EVGVGGKGA 540
GGVSANPALN GSAGANGTAP TSGGNGGNG AGATPTVAGE NGGAGGNGGH GGSVNGGAG 600
GAGNGVAGT GLALNGGNG NGGIGGNGG AAGTGGDGGK GGNGGAGANG QDFSASANGA 660
NGGQGGNGGN GGIGGKGGDA FATFAKANG GAGGNGGNG VAGQGGAGGK GAIPAMKAT 720
60 GADGTAPTSG GDGNGGNGA SPTVAGGNG DGGKGGSGN VNGGNGGAG GNGAAGQAGT 780
PGPTSGDSGT SGTGGAGGN GGAGGAGGTL AGHGGNGGK GNGGQGGIGG AGERGADGAG 840
PNANGANGEN GSGGNGGDG GAGGNGGAG KAQAAGYTDG ATGTGGDGN GGDGGKAGDG 900
GAGENGLNSG AMLPGGGTVG NPGTGGNGN GGNAGVGGTG GKAGTGSITG LDGTDGITPN 960
GGNGGNGGNG GKGGTAGNGS GAAGGNGNG GSGLNGGDG NGNGGGALN QAGFFGTGGK 1020
65 GGNGGNGGAG MINGGLGGFG GAGGGGAVDV AATTGGAGGN GCAGGFASTG LGGPGGAGGP 1080
GGAGDFASGV GGVGGAGGDG GAGGVGGFG QGGIGGEGRT GGNGGSGGDG GGGISLGGNG 1140
GLGGNGGVSE TGFGGAGGNG GYGGPGGPEG NGGLGGNGGA GGNGGVSTTG GDGGAGGKGG 1200

NGGDGCVGL GGDAGSGGAG GNGGIGTDAG GAGGAGGAGG NGGSSKSTTT GNAGSGGAGG 1260
NGGTGLNGAG GAGGAGGNAG VAGVSFGNAV GGDGNGGNG GHGGDGTGG AGGKGGNGSS 1320
GAASGSGVVN VTAGHGGNGG NGGNGGNGSA GAGGQGGAGG SAGNGGHGGG ATGGDGGNGG 1380
NGGNSGNSTG VAGLAGGAAG AGGNGGGTSS AAGHGGSGGS GSGTGGAG AAGGNGGAGA 1440
5 GGGSLSTGQS GGP RRQRWCR WQRRRWLGRQ RRRRWCRWQR RCRRQRWRWR CRQRLRRQW 1500
RQRRRCRPW LHRRRGRQGR RWRQRRFQQR QRSRWQRR 1538
<212> Type : PRT
<211> Length : 1538
SequenceName : SEQ ID 146
10 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
15 <400> PreSequenceString :
MSFVVTAPPV LASAASDLGG IASMISEANA MAAVRTTALA PAAADEVSAA IAALFSSYAR 60
DYQTLVSVQVT AFHVQFAQTL TNAGQLYAVV DVGNGVLLKT EQQVLGVINA PTQTLVGRPL 120
IGDGTGTHGAPG TGQNGGAGGI LWGNGGNGGS GAPGQPGGRG GDAGLFGHGG HGGVGGPGIA 180
GAAGTAGLPG GNGANGGSGG IGGAGGAGGN GLLFGNGGA GGQGGSGGLG GSGGTGGAGM 240
20 AAGPAGGTGG IGGIGGIGGA GGVGGHGSAL FGHGGINGDG GTGGMGGQGG AGGNGWAAEG 300
ITVIGIGEQQG QGGDGGAGGA GGIGGSAGGI GGSQAGGHG GDGGQGGAGG SGGVGGGGAG 360
AGGDGGAGGI GGTGGNGSIG GAAGNGGNGG RGGAGGMATA GSDGGNGGGG GNGGVGVGSA 420
GGAGGTGGDG GAAGAGGAPG HG YFQPPAPQ GLPIGTGGTG GEGGAGGAGG DGGQGDIGFD 480
GGRGGDGGPG GGGGAGGDGS GTFNAQANNG GDGAGGVGG AGGTGGTGGV GADGGRGDS 540
25 GRGGDGGNAG HGGAAQFSGR GAYGEGGSG GAGGNAGGAG TGGTAGSGGA GFGGNGADG 600
GNGGNGGNGG FGGINGTFGT NGAGGTGGLG TLLGGHNGNI GLNGATGGIG STTLTNATVP 660
LQLVNTTEPV VFISLNGGQM VPVLLDTGST GLVMSQFLT QNFGPVGITG TAGYAGGLTY 720
NYNTYSTTVD FGNLLTLPT SVNVTSSSP GTLGNFLSRS GAVGVLGIGP NNGFPGTSSI 780
VTAMPGLLNN GVLIDESAGI LQFGPNTLTG GITISGAPIS TVAVQIDNGP LQAPVMFDS 840
30 GGINGTIPSA LASLPSGGFV PAGTTISVYT SDGQTLTLYS TTTATNTPFV TSGGVMNTGH 900
VPFAQQPIYV SYSPTAIGTT TFN 923
<212> Type : PRT
<211> Length : 923
SequenceName : SEQ ID 147
35 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
40 <400> PreSequenceString :
MIGNGGAGGS GAPGAIGGAG GPAGLIGVGG AGGAGGDSAV AGVIGGAGGA GGAALLFGAG 60
GAGGAGGSGG SGAAGGAGGA GGAGGLFASG GSGGFGGFAS TGTGGAGGTG GAGGLFASGG 120
VGGTGGGAGS GGTGGVGGTG GAGGLFASGG AGGAGGSGGT GGAGGTGGAG GLFGAGGAGG 180
LGGQGNHTGG HGGAGGSAGL LALGDGGAGG AGGAATTGTG GAGGAGGKAG LLFGSGGAGG 240
45 SGGAAGTFGD TGNSGGAGGA GKGALLFGS GGAGGSGGAG GFANGSTGGA GGAGGGAGLI 300
GNGGNGGSGG TSVATGGAGN GGAGGAGGGA GLIGNGGNGG SGGMGDAPGG TGVGGIGGLL 360
LGLDGANAPA STNPLHTAQQ QALAAVNAPI QAVTGRPLIG NGANGAPGSG APGGHGGWLF 420
GGGGTGGSGV SGGAGGDGGA GGILFGAGGA GGAGGAVTGT GATGSGGAG GGAALLFGAGG 480
AGGAGGSSGI GGFAAGGAGG PGGAGGLFNG GGAGGAGGSG VSGGAGGEGG AGGAGGLFAG 540
50 GGAGGAGGSG NNVGGAGGAG GVGGLFGAGG AGGSGGGGSGV AGDSGAGGNA GLLAPGLAGG 600
AGGGGGQGFQD TGGAGGPGGD AGLLVGSGGV GGAGGFGLTT GPGGAAGGDA GLLFGSGGAG 660
GAGGSGRTDL GGAGGAGGKA GLIGNGGNGG AGGAGGNGGG DGGPGGAAGF LGNGGNGGNG 720
GTGTSAGSPG AGGAGGSLIG AEGLPGLLP 749
<212> Type : PRT
55 <211> Length : 749
SequenceName : SEQ ID 148
SequenceDescription :

Sequence

60 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIAAPEA LVAVASDLAG IGSALAEANA AALAPTTALL AAGADEVSAA IAALFGAHGQ 60
AYQTVSAQAS AFHAQFVQAL TGGGGAYAAA EAANVSAAQS TDQRLDLIN GPTQALLGRP 120
65 LIGDGANGGP GQDGGPGGLL YGNGGNGGTS TTAGVAGGNG GAAGLIGNGG AGGGGGAGAA 180
GGNGGAGGWL YGNGGAGGAG GTSVIPGVAG GNGGAGGSAG LWGTGGAGGD GNGRSGPVN 240
VAGSAGGNGG AGGAAGLFGD AGAGGNGGKG GAGGAAFSIN FTAGDGGAGG AGGSGGHALL 300

WGAGGAGGNG GSGGTGGAGG STAGAGGNGG AGGGGGTGGL LFGNGGAGGH GAAAGNGLAA 360
GNGVSSSGGG GAGGTGGAGG DGGAGGAGGN ARLWGVGGAG GAGGDGGAGG AGGKGGSGLS 420
GNANGGAGGD SGRGGTGGAG GEGGAAGLLV GTGGHGGDGG AGGAAVKGGD GGAAAGTGIA 480
GAGGRGGAGG SGGSGGDGGG GAAGPAGWLF GDGGAGGNGG AAAAGGAGGQ AGGGGGNGGN 540
5 GNGGNGGNG GNGATGGWLY GNGGAGGQGA TAGAGGAGAN GVSSTNGGGT GNGGGIGGTG 600
GSGGAGGNAG LLGVGGAGGH GASGGAGDRG GAGGTGFISS DGGAGGDGGD GNGGAGGTG 660
GLLFGAGGNG GPGGSGGAAD IGGNGGAGNG GGTGNGGNG GSGGGAGSGG DGGGAGGNGA 720
WLFNGGAGG GGGKGGNGAG GGLGGGSFGL PGLNGSGGDG GDGNGAPGG VLYGNGGAGG 780
QGSSGGIGGP GATGGAGGKG GDGGDAQLIG DGGNGGNGGA GGTGCTPGPG GPGGSGGLGG 840
10 LLFGQTGTAG VSP 853
<212> Type : PRT
<211> Length : 853
SequenceName : SEQ ID 149
SequenceDescription :
15
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
20 MSYLVVVPEL VAAAATDLAN IGSSISAANA AAAAPTALV AAGGDEVSA IAAALFGAHAR 60
AYQALSAQAA MFHEQFVRAL AAGGNSYAVA EAATAQSVQQ DLLNLINAPT QALLGRPLIG 120
NGANGLPGTG QNGGDGGILY GNGGNGGSGG VNQAGGNGGN AGLWNGGSGG GAGGNATTAG 180
RNGFNGGAGG SGGLLWNGG AGGAGGNGGP APLVGGVGTG GGAGGNGGGA GLFYGFGGAG 240
GNGGMGGVAP STGPSMGILP AGGVGGPGGS GGASALAFGS GGVGGAGGLG GPTDGTVQGV 300
25 GGFQGGQGGNG GQSGLLFGNA GAGGAGAAGG AGTGDTEFSG GHGGAGGDGG AVGLIGNGGA 360
GGTGSPGAVV GNGGVVGGLG GAGSPGGLLY GTGGAGGNGG PGGDGGTGAT VGFAGSGGFG 420
GAGGIAQLFG TGGMGGSGGG IGAGTTTVVP PDVAPVGGTG GNGGRAGLLL GVGGMGGNGG 480
ATSVGGTLYA AGNGGGDGL VWGNGGTGGS GGAGGAGSVG NGGAGGNAAL LFGNGGAGGA 540
GGAGGIGAGG AGGFGAVLFG NGGAGGSGAP GGIGAGGNGG NALLVNGGN GGAGTGAAG 600
30 GAGGSGGLLF GQNGMPGP 618
<212> Type : PRT
<211> Length : 618
SequenceName : SEQ ID 150
SequenceDescription :
35
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
40 MNFSVLPPET NSALIFAGAG PEPMAAATA WDGLAMELAS AAASFGSVTS GLVGGAWQGA 60
SSSAMAAAAA PYAAWLAAAA VQAEQTAAQA AAMIAEFEAV KTAVVQPM LV AANRADLVSL 120
VMSNLFQNA PAIAAIEATY EQMWAADVSA MSAYHAGASA IASALSPFSK PLQNLAGLPA 180
WLASGAPAAA MTAAAGIPAL AGGPTAINLG IANVGGGNVG NANNGLANIG NANLGNYNFG 240
SGNFGNSNIG SASLGNNGN FGNLGSNNVG VGNLGNLNTG FANTGLGNFG FGNTGNNNIG 300
45 IGLTGNNQIG IGGLNSGTGN FGLFNSGSGN VGFFNSGNGN FGIGNSGNFN TGGWNSGHGN 360
TGFFNAGSFN TGMLDVGNAN TGSLNTGSYN MGDENPGSSN TGTFNTGNAN TGFLNAGNIN 420
TGTFNIGHMN NGLFNTGDMN NGVFYRGVGO GSLQFSITTP DLTLPPLQIP GISVPAFSLP 480
AITLPSLNIP AATTPANITV GAFSLPGLTL PSLNIPAAAT PANITVGAFA LPGLTLPSLN 540
IPAATTPANI TVGAFLPGL TLPSLNIPAA TTPANITVGA FSLPGLTLPS LNIPAAATTPA 600
50 NITVGAFLP GLTLPSLNIP AATTPANITV SGFQLPPLSI PSVAIPPVTV PPITVGAFLN 660
PPLQIPEVTI PQLTIPAGIT IGGFSLPAIH TQPIVQGIG VGQFGLPSIG WDVFLSTPRI 720
TVPAFGIPFT LQFQTNVPAL QPPGGGLSTF TNGALIFGEF DLPQLVHPY TLTGPIVIGS 780
FFLPANIPG IDVPAINVDG FTLPQITTPA ITTPEFAIPP IGVGFTLPQ ITTQEIITPE 840
LTINSIGVGG FTLPQITTP ITTPPLTIDP INLTGFTLPQ ITTPPITTP LTIDPINLTG 900
55 FTLPQITTP ITTPPLTIEP IGVGFTTPP LTVPGIHLPS TTIGAFAPG GPGYFNSSTA 960
PSSGFFNSGA GGNSGFGNNG SGLSGWFNTN PAGLLGSGY QNFGGLSSGF SNLGSVSGF 1020
ANRGILPFSV ASVVS GFANI GTNLAGFFQG TTS 1053
<212> Type : PRT
<211> Length : 1053
SequenceName : SEQ ID 151
SequenceDescription :
Sequence

65 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MLYVVASPDL MTAAATNLAE IGSAISTANG AAALPTVEVV AAAADEVSTQ IAAALFGAHAR 60

SYQTLSTQAA AFHSRFVQAL TTAAASYASV EAANASPLQV ALDVINAPAQ TLLGRPLIGN 120
GADGSTPGQA GGPGLLYGN GGNGAAGGPN QAGGAGGNAG LIGNGGAGGA GVGAVGGKR 180
GTGGLLFGNG GAGGQGLGL AGINGGSGGQ GGHGGNAILF GQGGAGGPGG TGAMGVAGTN 240
PTPIGTAAAPG SDGVNQIGNG GNTDLTGGAG GDGNAGSTTV NGGNGGTGGA ARNSSGGTGN 300
5 SFGGAGGAGG DGANGGDGGA GGEALTEGGA TAVSGAGGKG GNAEASGGAG GNGGKGGFAQ 360
ATTSVTGGNG GNGGNHDSN APGGAGGSGG VGGDGGRGGL LAGNGGTGGA GNGGTGGAG 420
APGGAGGAGG KADIANSIGD NATVTGGNGG TGGDGGSALG TGGAGGAGGL GGHGGAGGLL 480
IGNGGAGGAG GLGGAGGAGG AGGEGGAGGA GGEAIPGGAS TNSAGGDGGA GGTGGNGGDG 540
GAGGAPGLGG AGGAGGWLIG QSGSTGGGGA GGAGGAGGAG GAGGSGGAGG HGDTTSGKNG 600
10 SSGTAGFDGN PGQPG 615
<212> Type : PRT
<211> Length : 615
SequenceName : SEQ ID 152
SequenceDescription :
15
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
20 MHYSVLPPEI NSALIFAGAG SGPMIAAASA WDGLATELAS AAVSFGSVTA GLVGGSWQGR 60
SSVAMAAAAA PYAGWLAAAA TQAEQAATQA QVMVAEFEAV RLAMVQPALV AANRSLISL 120
VISNLFQNA PAIAAAEAAY EEMWALDVSA MAAYHSGASA VAVALPAFAL PLRLPAGLAA 180
GPAAVVLTAL TAVGMPFTAG RAIAASLGLA NVGGGNLNA NNGLGNIGNA NLGNNNLGSG 240
NFGSFNIGSA NLGGNNIGIG NAGANNFGLA NLGNLNTGFA NAGIGNFGIA NTGNNNIGNG 300
25 LTGNNQIGIG GLNSGNGNVG LFNAGSANIG FFNSGNGNFG IGNSGNFSTG LFNPGHGNTG 360
FLNAGSFNTG MFDVGNANTG SFNVGHYNFG AFNPGPSNTG TFNTGGANTG WFNTGSINTG 420
AFNIGDMNNG LFNTGDMNNG VFYRGVGGQS LQFAITSPDL TLPSLEIPGI SVPAFSLPAI 480
TLPSLTIPAV TTPANVTVGA FDLPLGLTVPS LTIPAAATPA NITVGAFDLP GLTVPSLTIP 540
ATTTANITV GAFNLPQLSI PSVTVPPI TI PAGTALGAFN LPTLSIPSVT VPPITIPAGT 600
30 TVGGFTLPTI HTPLISTPQI SIGGFSTPGI ATQANSVIN LPTFSLNGIT ITNLVVFIPN 660
NITALQTNMP GVFPQIGGFA NTPPAFINTG TITVGGGQIN GVGFSGAIN VTPFTLPNVV 720
IQPWSLGGIS VDGFTLPEIS TQFTTTPALT ISPIGVGALS LPDITTOQFT TPELTIDPIT 780
LGGFTLPQLS IPAITTPAFT IDPIALGGFT LPQIMTPEIT TPPFAIDPIG LSGFTLPQVN 840
IPEITTPFT IQPVGLAAFT TPALTIASIH LPSTTMGGFA IPAGPGYFNS SATPSLGFFN 900
35 AGIGGNSGFG NSGSLSGWF NTSPVGLLAG SGYQNYGGLI SGFSNLGSGI SGFANTGTLP 960
FAVTSLSVGL ANIGNNLSGL FFQSTTP 987
<212> Type : PRT
<211> Length : 987
SequenceName : SEQ ID 153
SequenceDescription :
40
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
45 MSFVVVAPEV LAAAASDLAG IGSTLAQANA AALAPTTAVL AAGADEVSAA IASLFGAHGQ 60
AYQAVSAQMS AFHAQFMQAL TGAGGAYAAA EAVNVSAQVS VEQDLIAAIN ARFERIFGRP 120
LIGDGANGGP GQDGGPGGLL YGNGGNGGTS TTVGMAGGNG GAAGLIGNGG FGGGGGPGAA 180
GGNGGAGGWL FGNGGAGGAG GLGVAPGVPG GAGGAGGAGG VGGPAGLWGH GGAGGAGGAG 240
50 VAGAGGFEGT IGAGGAGGVG GAGGVGGAGG AGGWLYGDAG AGGDGGVGGG GGTGGLGNRG 300
GAGGAGGAGG VGGAGGAAGL WGGGGAGGVG GTGGGAGLGA QSVTFSSSLG GLSGGDGGAG 360
GAGGAGGAGG TGGWLYGGGG AAGSGGDGGT GGQGGAGGAG VFSLFSGGGG PGGNGGVGGV 420
GGVGGAGGRA GLFGVGGLGG AGGDAGDSGE GFGGPGLAG GLFGNPGNGG VGGIGGDAAA 480
GGAGGAGGNG GAGGNGWLF GNGGAGGSGG DGAAGRGGA GNLSAGGIN APAGNPGSGS 540
55 VGIGGAGGAG GTAGLFGDGG AGGAGGAGAA GFGGGISAAT PSAGSEGAMG GAGGVGNAR 600
LLGTGGAGGV GGGGAGGDG GRGGVATPGG QGGDAGDGGA GGAGGNGGGA SGAGGWLLGT 660
GGAGGAGGNG GNGGKAGFSP GPTNFGLNGA GGGGGVGGNG ATGPWLFGDG GPTPGSTGAG 720
AAGGHGGDAQ LIGNGGHGA GGTGVPNGSG GAGGLSGLLF GEPGANG 767
<212> Type : PRT
60 <211> Length : 767
SequenceName : SEQ ID 154
SequenceDescription :
Sequence

65 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

MSFVIANPEM LAAAATDLAG IRSAlSAATA AAAAPTlQVA AAGADEVSLA ISALFGQHAQ 60
AYQALSAQAT IFHDQFVQAL TSGGNLYAAA ESHTVEQMVL NAINAPTQTL FGRPLIGDGA 120
NGTAENPDGQ NGGLLFGNGG NGFTQTTAGV AGGNGGSAGL IGNGGAGGGG GAGAAGGLGG 180
NGGWLYGNGG AGGIGGAGTG TGGHGGAGGA GGRWLWGTG GAGGAGGDGG WLFGDGGAGG 240
5 TGGNGGSGFN SLTSSVGGAG GAGGHAGLFG AGGTGGTGGI GGQNTETGPA ASNGGAGGAG 300
GGGGYLVGDG GAGGTGGAGG KNSSGGATLT GGTGGTGGAG GAAGWLYGSG GAGGAGGAGG 360
LNNAGGATGG TGGTGGAGGS GAWLYGNNGA AGAGNGNGNN TSAGTGGVGA SGGTGGNAGL 420
IGAGGHGGAG GAGGNQTGGV GNGGAGGNGG AGGAGGQLYG NGGDGGNGGA GGANIAGGNG 480
SDGGAAGHGG AGGSARLIGA GGHGGDGGAG GNTAGRRADA IAGTGGDGGN GGNGGLLSGN 540
10 AGAGGHGGAG GSSTATTTTG TPPTGATGGN GGNGGAGGTA GFTGSGGIGG NGGAGGTGGN 600
AGVALSVGST GGLGGNGGSG GLGGGGGSLF GNGGAGGVGA TGGNGSGIG PASVGGNGGK 660
GGVGAAGGLA GQIGNGSGG SGGAGNGGT GDTAGNGNG GAGAVGGNAQ LIGNGGNGG 720
GGNGGTGADG T 731
<212> Type : PRT
15 <211> Length : 731
SequenceName : SEQ ID 155
SequenceDescription :

Sequence
20 -----
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MPGRFRNFGS QNLGSGNIGS TNVSGSNIGS TNVSGSNIGD TNFGNGNNGN FNFGSGNTGS 60
NNIGFGNTGS GNFGFGNTGN NNIGIGLTGD GQIGIGLNS GSGNIGFGNS GTGNVGLFNS 120
25 GTGNVGFNS GTANTGFGNA GNVNTGFWNG GSTNTGLANA GAGNTGFFDA GNYNFGSLNA 180
GNINSSFGNS GDGNSGFLNA GDVNSGVNA GDVNTGLNS GNINTGGFNP GTLNTGFFSA 240
MTQAGPNSGF FNAGTGNSTG GHNDPAGSGN SGIQNSGFGN SGYVNTSTTS MFGGNSGVLN 300
TGYGNSGFYN AAVNNTGIFV TGVMSGFFN FGTGNSGLLV SGNLSGFFK NLFG 354

30 <212> Type : PRT
<211> Length : 354
SequenceName : SEQ ID 156
SequenceDescription :

Sequence
35 -----
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVLAMPEV LGSAAIDLAA LGSVLGAADA AAAATTTGIV AAAQDEVSA AALFSAHGR 60
40 AYQVASAQAA AVHAQFVEAL SAGAGAYASA EAAGAAVLAN PAQSVQDILL AAVNAQSVAL 120
TGRPLIGNGA NGAPGTGANG APGGWLLGNG GAGGSAAAGS GLPGGAGGAA GLFGTGGAGG 180
AGGSSTVGDG EAGGAGGSGG WLLGTGGVGG VGGLGAGAGG AGGVGGAGGL LGAGGHGGAG 240
GLGAVTGGVG GTGGAGGLLA GLLAGPGGAG GTGGRGFLNN GGVGAGGNA GLLFGAGGTG 300
GSGGAGLGGD GGAGGAGGNT GVLFGNAGSG GTGGFGDTDG GAGGAGGDAG WLGS GGVGGA 360
45 GGFGETGDGG VGGAGGKAGL LIGNGGAGGA GGQGA VTGGT GGAGGDGVL GNGGNAGIGG 420
TGPTAGDTGA GGISGLLLGA DGFNTPASAS PLHTLKQAL AAINAPTQTL TGRPLIGNGT 480
PGAVGSGATG APGGWLLGDG GAGGSGAAGS GAPGGAGGAA GLWGTGGAGG AGGSAGGGG 540
AGGAGGAGGW LLGDGGAGGI GGASTVLGGT GGGGGVGLW GAGGAGGAGG TGLVGGDGA 600
GGAGGTGGLL AGLIGAGGH GGTGGLSTNG DGGVGAGGN AGMLAGPGA GGAGGDGENL 660
50 DTGGDGGAGG SAGLLFGSGG AGGAGGFGFL GDDGGAGGNA GLLLSGGAG GFGGFGTAGG 720
VGGAGGNAGW LGFGGAGGVG GSAGLIGTGG NGNGGTGAN AGSPGTGGAG GLLLGQNGLN 780
GLP 783
<212> Type : PRT
<211> Length : 783
55 SequenceName : SEQ ID 157
SequenceDescription :

Sequence
60 -----
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSLVIATPQL LATAALDLAS IGSQVSAANA AAAMPTTEVV AAAADEVSAA IAGLFGAHAR 60
QYQALSQVA AFHEQFVQAL TAAAGRYAST EAAVERSLLG AVNAPTEALL GRPLIGNGAD 120
GTAPGQPGA GLLFGNGGN GAAGGFGQTG GSGGAAGLIG NGNGGAGGT GAAGGAGGNG 180
65 GWLWNGGNG GVGTSVAAG IGGAGNGGN AGLFGHGAG GTGGAGLAGA NGVNPTPGA 240
ASTGDSADV SGIGDQTGGD GGTGGHGTAG TPTGGTGGD ATATAGSGKA TGGAGDGGT 300
AAAGGGGNG GDGGVAQDI ASAFGGDGGN GSDGVAAGSG GSGGAGGA FVHIATATST 360

GGSGGFGGNG AASAASGADG GAGGAGGNNG AGGLLFGDGG NGGAGGAGGI GGDGATGGPG 420
GSGGNAGIAR FDSPDPEAEP DVVGGKGGDG GKGGSGLGVG GAGGTGGAGG NGGAGGLLFG 480
NGGNNGNAGA GGDGGAGVAG GVGGNGGGGG TATFHEDPVA GVWAVGGVGG DGGSGGSSLG 540
VGGVGGAGGV GKGGASGML IGNGGNNGSG GVGGAGGVGG AGGDGGNNGS GGNASTFGDE 600
5 NSIGGAGCTG GNNGNGANGG NGGAGGIAGG AGGSGGFLSG AAGVSGADGI GGAGGAGGAG 660
GAGSGGGEAG AGGLTNGPGS PGVSGTEGMA GAPG 694
<212> Type : PRT
<211> Length : 694
SequenceName : SEQ ID 158
10 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
15 <400> PreSequenceString :
MSFVIAAPEV IAAAATDLAS LESSIAAANA AAAANTTALL AAGADEVSTA VAALFGAHGQ 60
AYQALSAQAQ AFHAQFVQAL TSGGGAYAAA EAAATSPLLA PINEFFLANT GRPLIGNGTN 120
GAPGTGANGG DGGWLIGNGG AGGSGAAGVN GGAGNGGAGG GLIGNGGAGG AGGRASTGTG 180
GAGGAGGAAG MLFGAAGVGG PGGFAAAFGA TGGAGGAGGN GGLFADGGVG GAGGATDAGT 240
20 GGAGGSGGNG GLFGAGGTGG PGGFGIFGGG AGGDGGSGGL FGAGGTGGSG GTSIINVGGN 300
GGAGGDAGML SLGAAGGAGG SGGSNPDGGG GAGGIGGDGG TLFGSGGAGG VCGLGFDAGG 360
AGGAGGKAGL LIGAGGAGGA GGGSFAGAGG TGGAGGAPGL VGNAGNGGNG GASANGAGAA 420
GGAGGSGVLI GNNGNNGSGG TGAPAGTAGA GGLGGQLLGR DGFNAPASTP LHTLQQQILN 480
AINEPTQALT GRPLIGNGAN GTPGTGADGG AGGWLFNGG NGGHGATGAD GGDGGSGGAG 540
25 GILSGIGGTG GSGGIGTTGQ GGTGGTGGAA LLIGSGGTGG SGGFGLDTGG AGGRGDDAGL 600
FLGAAGTGGQ AALSQNFIGA GGTAGAGGTG GLFANGGAGG AGGFGANGGT GGNLLFGAG 660
GTGGAGTLGA DGGAGGHGGL FGAGGTGGAG GSSGTFGGN GSGGNAGLL ALGASGGAGG 720
SGGSALNVGG TGGVGGNNGS GGSFLFGGGA GGTGGSSGIG SSGTGGDGG TAGVFGNGGD 780
GGAGGFGADT GGNSSSVRNA VLLIGNGGNG NGGKAGGTPG AGGTSGLIIG ENGLNGL 837
30 <212> Type : PRT
<211> Length : 837
SequenceName : SEQ ID 159
SequenceDescription :

35 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
40 MSFVIAVPET IAAAATDLAD LGSTIAGANA AAAANTTSLI AAGADEISAA IAALFGAHGR 60
AYQAASAEAA AFHGRFVQAL TTGGGAYAAA EAAAVTPLLN SINAPVLAAT GRPLIGNGAN 120
GAPGTGANGG DAGWLIGNGG AGGSGAKGAN GGAGGPGGAA GLFGNGGAGG AGGTATANNG 180
IGGAGGAGGS AMLFGAGGAG GAGGAATSLV GGIGGTGGTG GNAGMLAGAA GAGGAGGFSF 240
STAGGAGGAG GAGGLFTTGG VGGAGGQGHT GGAGGAGGAG GLFGAGGMGG AGGFGDHGTL 300
45 GTGGAGGDGG GGGLFGAGGD GGAGGSLTGT GGAAGNGGNA GTLSLGAAGG AGGTGGAGGT 360
VFGGGKKGAG GAGGNAGMLF GSGGGGGTGG FGFAAGGQGG VGGGAGMLSG SGGSGGAGGS 420
GGPAGTAAGG AGGAGGAPGL IGNGGNNGNG GESGGTGGVG GAGGNAVLI NGEGEGIGAL 480
AGKSGFGGFG GLLLGADGYN APESTSPWHN LQQDILSFIN EPTEALTGRP LIGNGDSGTP 540
GTGDDGGAGG WLFNGGNGG AGAAGTNGSA GGAGGAGGIL FGTGGAGGAG GVGTAAGAGG 600
50 GGAGGSAFLI GSGGTGGVGG AATTTGGVGG AGGNAGLLIG AAGLGGCGGG AFTAGVTTGG 660
AGGTGGAAGL FANGGAGGAG GTGSTAGGAG GAGGAGGLYA HGGTGGPCGN GGSTGAGGTG 720
GAGGPGLYAG AGGSGGAGGH GGMAGGGGGV GGNAGSLTLN ASGGAGGSGG SSLSGKAGAG 780
GAGGSAGLFY GSGGAGGNGG YSLNGTGGDG GTGGAGQITG LRSFGGAGG AGGASDTGAG 840
GNGGAGGKAG LYGNNGDGGA GGDGATSGKG GAGGNAVIG NGNGGNAGK AGGTAGAGGA 900
55 GGLVLGRDQ HGLT 914
<212> Type : PRT
<211> Length : 914
SequenceName : SEQ ID 160
SequenceDescription :

60 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
65 MSLVIVTPET VAAAASDVAR IGSSIGVANS AAAGSTTSVL AAGADEVSAA IATLFGSHAR 60
EYQAISTQVA AFHDRFAQTL SAAVGSYVSA EATNAAPLAT LEHNVNALN APTQALLGRP 120
LIGDGAAGAP GTGQAGGAGG ILWNGGAGG SGAPGQVGA GGAAGLFGTG GAGGAGGAGA 180

AGGAGCSGGW LLGNGGVGGA GGQSLGGAT GGAGGNAGLF GVGGTGGPGG PGGPGGVGGT 240
GGAGGLGGTL YGAGGHGGAG GPGPIGGVGG HGGVGAAGL LGVGGHGGAG GHGAEGVAGA 300
AGEDLSPHGT SGGVGGDAGD GGTGGRGGWL AGAGGAGCAG GVGGTGGAGG AGFSRALIVA 360
GDNGGDGGNG GMGGAGGAGG PGGAGGLISL LGGQAGGAG GTGGAGGVGG DRGAGGPGNQ 420
5 AFNAGAGGAG GHGGDPGAGG AGGTGGAGSI TGAQGAIGAT PTSGGNGGAG GNGANATTAG 480
TNGANGGPGG HGGLVGNNGA GGNGANGAAG TNASDSGAVG GKGNSGGNGG QGGAGGDGGT 540
LAGNGGAGGT GGRGADGGLG GSGAEGANAT TAGERGQDGG KGGNGGVGGT GGNVAVPGAN 600
GGHGGNGGNP GFSGAGGLGG LSGDGVTRAA QGATPDFADT GKGNGGNGG ANAVAPGGTG 660
ASGGAGGNAG AGGKGGENTII GDGGGGNGGA GKGAGGATLL GLTVFGDNGG AGVLGDSTDP 720
10 DSGGAGGAG GAGGAGDPT I 741
<212> Type : PRT
<211> Length : 741
SequenceName : SEQ ID 161
SequenceDescription :
15 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
20 MSFVTAAPEM LATAAQNVA IGTSLSAANA TAAASTTSVL AAGADEVSQA IARLFSDYAT 60
HYQSLNAQAA AFHHSFVQTL NAAGGAYSSA EAANASALQ EQNLLAVINA PAQALFGRPL 120
IGNGANGTAA SPNGGDGGIL YGNGGNGFSQ TTAGVAGGAG GSAGLIGNGG NGGAGGAGAA 180
GGAGGAGGWL LGNGGAGGPG GPTDVPAGTG GAGGAGGDAP LIGWGGNGGP GGFAAFNGGG 240
AGGNGGASGS LFGVGGAGGV GGSSDEVGGT GGAGGAGRGL FLGLGGDGGG GGTSNNNGGD 300
25 GGAGGTAGGR LFSLGGDGGN GGAGTAIGSN AGDGGAGGDS SALIGYAQGG SGGGIGFGES 360
TGGDGGGLGA GAVLIGTVVG GFGGLGGSN GTGGAGGAGG TGATLIGLGA GGGGIGGFA 420
VNVGNGVGGG GGQGGQAAL IGLGAGGAGG AGGATVVGLG GNGGDGGDGG GLFSIGVGGD 480
GGNAGNGAMP ANGGNGGNAG VIANGSFAPS FVGFGNGGN GVNGGTGGSG GILFGANGAN 540
GPS 543
30 <212> Type : PRT
<211> Length : 543
SequenceName : SEQ ID 162
SequenceDescription :
35 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
40 MSYVLATPEM VAAAANNLAQ IGSTLSAANA AALAPTTGVL AAGADEVSAA VASLFSGHAQ 60
AYQTLGTQAA AFHERFIQAL STAAGAYGSA EAANASPLQQ ALNVINAPTQ TLLGRPLIGN 120
GTNGAPGTGQ AGGPGGLLYG NGGNGSGGV GQAGGAGGSA GLIGIGGTGG AGGAGAVGGV 180
GGNGGWLYGN GGAGGLGGTG VAGVNGMGMA AGGAGGNAYL FGSGGAGGQG GMGAAGADGV 240
NPTPTGTADA GSTGTDQTLG GNAIGNGGP GDAGDAMTSG GAGGSGGNAV STVNGDAVGG 300
EGGKGEGAY GGAGGAGGSA ASIGNAAIGG NGGAGGNAQA PGGVGGAGGE GGDAQVGTNS 360
45 PSNAEAGNGG SGGNGFDSFA SGGTGGAGGT GGAGGRGGLL IGDGAGGAG GVGGTGGSGA 420
PGGGGAGGD GGAANTDSAG SSRKAFGGDG GVGGDGASAL GTGEGGIGG QGGNGGAGGL 480
LIGNGAGGV GGTAGAGTG GSGGAGGAG AGGGGTNSGP GAAFGNGNT GGNGNGGAP 540
GALGGKGGSG GLIGRAGSDG GVGAGGAGGA GGAGGTGGEG GTGGDGKTTD GNPGMGGSPG 600
SAGQPG 606
50 <212> Type : PRT
<211> Length : 606
SequenceName : SEQ ID 163
SequenceDescription :
55 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
60 MSFLFAQPEM LGAAATDLAS IGSAISTANA AAAAATTRVL AAGADEVSAA VAALFSGHAQ 60
TYQALRTQAA AFHQIIVQTL TSTAGAYASA EAANVEQQLL GAINAPTMAL LGRPLIGHGA 120
DGAPGTGQAG GAGGILYNGG NGGSGATGQ AGGAGGAAGL IGHGGAGGLG GTGASGGAGG 180
AGGWLWNGG AGGNGGVGVA GDPGGVGGAG GAGGAAGLWG SGGSGGTGGQ GGVGGKSGD 240
GGTGGIGGAG GGGWLHGDG GAGGHGGQGG TGVSSGGNGG AGGTGGDGRG LSGSGGAGGR 300
GGQTGVGGKV GENNFAGG AGGTGGLIGN GGAGNGGQG AISGAGGAGG NAWLIGDGA 360
65 GGNGGDIRGQ GGGAGGAGGA GGQLIGNGGT GGAGGTVTSP NGLGGAGGAG GSAGLIGHGG 420
TGGAGGHSAG GPDGNGGIGG AGGAGGNGGQ LYGTGGTGGT GKGGDGFGV FGKGGAGGTG 480
GRGGAAGLIG DAGTGGTGK GGTAGEDGTG NGGTTGGNGG AAVLIGNGGG GGAGNGGAG 540

NDGTPGNNGG GGVGGTGGTL FGQPGQPGPP GQPGPA 576
<212> Type : PRT
<211> Length : 576
SequenceName : SEQ ID 164
5 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
10 <400> PreSequenceString :
MWTSQMIVAP AFVDAAAKDL ATIGSAISRA NAEALVPITA LLPAGADDVS AAIAALFATH 60
GQAYQELSAH AVAFHEQFVQ LMSAGAAQYA SAEANSSPL QIVGQTALDA INSPVQTLTG 120
RPLIGNGANG VAGTGQNGGD GGWLYGNNGN GSGGTGQNG GNGGSAGLWG SGGNGGQGGA 180
GANGAAGQPG KAGGSGGNGG AGGWIYGHGG HGGAGGNGGN ATAPCGASAG FDGGAGGNGG 240
15 SGGRGGLLFG NGGNGSVGGM GCQGTNDTAG DSAGSGGLGG NGNGAQGGW LIGNGGQGGD 300
SGAGGGTDST QTGVMNGASG GSAGIAGNGG DAGLVGNNGA GGNGNGAAG SALGTTIFGG 360
SGGVGGSGGD GNGGWLFGS GASGNGGQG GDAGTNGFAG FGGSAGGGGW VGAVNFGPIS 420
VQGFGLFGHG GDGNGGDVG AGSLSIQFGA SGGDGGQGGV LYNGGNGGN AGSGGTGFE 480
GSAGQGGAAL LIGNGGAGGN GATGGTGVGN IIQEAGGDGS DGGAGSGGGL LFGSGAGGI 540
20 GGAGGVGGSG NDGNGGDGG QGGASGLGIG NGGPGSGGT GGAGGTGSA GTGGAGDGG 600
NAALLIGTGG DGGDGVPPAP GGQGGKGLI GLPGQNGQP 639
<212> Type : PRT
<211> Length : 639
SequenceName : SEQ ID 165
25 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
30 <400> PreSequenceString :
MSWVMVSPDL VVAAAADLAG IGSAISSANA AAVNTTGLL TAGADEVSTA IAALFGAQGG 60
AYQAASAQAA AFYAQFVQAL SAGGGAYAAA EAAAVSPLA PINAQFVAAT GRPLIGNGAN 120
GAPGTGANGG PGGWLIGNGG AGGSGAPGAG AGGNGGAGGL FGSGGAGGAS TDVAGGAGGA 180
GGAGGNAGML FGAAGVGGVG GFSNNGATGG AGGAGGAGGL FGAGRERGS GSGNLTGGAG 240
35 GAGGNAGTLA TGDGGAGGTG GASRSGGFGG AGGAGGDAGM FFGSGSGGGA GGISKSVGDS 300
AAGGAGGAPG LIGNGGNGGN GASTGGGDG GPGGAGGTGV LIGNGGNGGS GGTGATLGKA 360
GIGGTGGVLL GLDGFAPAS TSPLHTLQD VINMVNDPFQ TLTGRPLIGN GANGTPGTGA 420
DGGAGWLFG NGGNGGQGTI GGVNGGAGGA GGAGGILFGT GGTGSGGPG ATGLGGIGGA 480
GGAALLFGSG GAGGSGGAGA VGGNGGAGGN AGALLGAAGA GGAGGAGAVG GNGGAGGNGG 540
40 LFANGGAGGP GFGSPAGAG GIGGAGGNGG LFGAGGTGGA GGGSTLAGGA GGAGGNGGLF 600
GAGGTGGAGS HSTAAGVSGG AGGAGGDAGL LSLGASGGAG GSGGSSLTAA GVVGGIGGAG 660
GLLFGSGGAG GSGGFSNSGN GGAGGAGGDA GLLVSGGAG GAGASATGAA TGGDGGAGGK 720
SGAFGLGGDG GAGGATGLSG AFHIGGKGV GSAVLIGNG GNGGNGGNSG NAGKSGGAPG 780
PSGAGGAGGL LLGENGLNGL M 801
45 <212> Type : PRT
<211> Length : 801
SequenceName : SEQ ID 166
SequenceDescription :

50 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
GQSYQAVSAQ AAFAHDFVQ LLNAGGGSYA SAEIANAQQN LLNAVNAPTQ TLLGRPLVGD 60
55 GADGASGPVG QPGGDGGILW GNGGNGGDST SPGVAGGAGG SAGLIGNGGR GGNGAPGGAG 120
GNGGLGGLLL GNGGAGGVGG TGDNGVGD LG AGGGGGDGL GGRAGLIGHG GAGGNGGDGG 180
HGGSGKAGGS GSGGFGQFG GAGGLLYGNG GAAGSGGNGG DAGTGVSSDG FAGLGGSGGR 240
GGDAGLIGVG GGGGNGGDP GLGARLFQVG SRGDDGGVGG WLYGDGGGGG DGGNGGLPFI 300
GSTNAGNGGS ARLIGNGGAG GSGSGAPGS VSSGGVGGAG NPGSGGNGG VWYNGGAGG 360
60 AAGQGGPGMN TTSPGGPGGV GGHGGTAILF GDGGAGGAGA AGGPCTPDGA AGPGSGGTG 420
GLLFGVPGPS GPDG 434
<212> Type : PRT
<211> Length : 434
SequenceName : SEQ ID 167
65 SequenceDescription :

Sequence

```
-----
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
5 MAHFSVLPPE INSLRMYLGA GSAPMLQAAA AWDGLAAELG TAASSFSSVT TGLTGQAWQG 60
  PASAAMAAAA APYAGFLTTA SAQAQLAAGQ AKAVASVFEA AKAAIVPPAA VAANREAFLA 120
  LIRSNWLGLN APWIAAVESL YEEYWAADVA AMTGYHAGAS QAAAQLPLPA GLQQFLNTLP 180
  NLGIGNQGNA NLGGGNTGSG NIGNGNKGSS NLGGGNIGNN NIGSGNRGSD NFGAGNVGTG 240
  NIGFGNQGP I DVNLLATPGQ NNVGLGNIGN NNMFGFNTGD ANTGGGNTGN GNIGGGNTGN 300
  NNFGFGNTGN NNIGIGLTGN NQMGINLAGL LNSGSGNIGI GNSGTNNIGL FNSGSGNIGV 360
10 FNTGANTLVP GDLNNLGVGN SGNANIGFGN AGVLNTGFGN ASILNTGLGN AGELNTGFGN 420
  AGFVNTGFDN SGNVNTGNGN SGNINTGSWN AGNVNTGFGI ITDSGLTNSG FGNTGTDVSG 480
  FFNTPTGPLA VDVSGFFNTA SGGTVINGQT SGIGNIGVPG TLFGSVRSGL NTGLFNMGTA 540
  ISGLFNLRLQ LG 552
  <212> Type : PRT
15 <211> Length : 552
      SequenceName : SEQ ID 168
      SequenceDescription :

Sequence
20 -----
  <213> OrganismName : Mycobacterium tuberculosis H37Rv
  <400> PreSequenceString :
  MSFLIASPEA LAATATYLTG IGSAISAANA VAAAPTTEIL AAGTDEVSTA ISALFGAHAQ 60
  AYQALSAHVA AFHDQFVHTL TAGAGSYMAA EAAAASPLQA LQLELLNAIN APTLALLGRP 120
25 LIGDGTDAAP GSGGAGGAGG ILIGNGGTGG ASDLAGTGRG GVGAGGAGG LFGIGGAGGG 180
  CGSAVAIGGD GGAGGAGGVF SGGGAGGAGD AIGGSGGAGG TGGLLGGGGG AGGAGGAGGN 240
  GGGASNSASI GGDGSGGAG GMLYGAGGVG GNGGAAVAIG GDGGAGGRAG AIGNGDGGN 300
  GGTSNTPGGS GGDGNGGNA GLIGNGGNGG NAEIVISGGS VAGTGGNGGL LLGFNGTNGL 360
  P 361
30 <212> Type : PRT
  <211> Length : 361
      SequenceName : SEQ ID 169
      SequenceDescription :

Sequence
35 -----
  <213> OrganismName : Mycobacterium tuberculosis H37Rv
  <400> PreSequenceString :
  AQASPAAHGG SGGAGGNGGA GSAGNGGAGG AGGNGGAGGN GGGGDAGNAG SGGNGGKGGD 60
40 GVGPGSTGGA GKGGAGANG GSSNGNARGG NAGNGGHGGA GSGDTGGAG GAGGQGGFGG 120
  TGGSGSGIGG GAGGNGGNGG AGGTGVVLGG KGGDGGNGDH GGPATNPGSG SRGGAGGSGG 180
  NGGAGGNATG SGGKGGAGGN GGDGSFGATS GPASIGVTGA PGGNGGKGA GGSNPNGSGG 240
  DGGKGGNGGA GNGGSGIGAN SGIVGSGGGA GGAGGAGGNG SLSSGEGGKG GDGGHGGDGV 300
  GGNSSVTQGG SGGGGGAGGA GSGGFFGGKG GFGGDGGQGG PNGGGTVGTV AGGGGNGGVG 360
45 GRGGDGVFAG AGGQGGGAGG GNGGGSTGG NGGLGGAGGG GGNAPDGGFG GNGGKGGQGG 420
  IGGGTQSATG LGGDGGDGGD GNGGNSGAK AGGAGGKGQA GQPNSGTEPG FGGDGGGLGA 480
  GATP 484
  <212> Type : PRT
  <211> Length : 484
50 SequenceName : SEQ ID 170
  SequenceDescription :

Sequence
55 -----
  <213> OrganismName : Rickettsia prowazekii
  <400> PreSequenceString :
  MKKSKILRKF LATASLCCTL FTNSNATGTI IPNNGSVSLN TDAGLVGGVF NNGDIIQIVN 60
  GGREIKISAD KANAIIIGIN TLKELPDFGG VEVSQNVSIG PLNAGEDLNT NFGPLKFISN 120
  NVTSIITGVG TKTFSNIDFA GKATLQINK DLNITTKIDN TVAGNNGSIT FEGSGIISNH 180
60 IGYTNSLLGI NVGNGEAKIY APEANNITIN AKNINLTHNN SILTLCGNI TTLKGNINNT 240
  TEIDGQGILN LAYDLGSSSI ITGDIGNIGS LDTINVLLGS ATFNSTILKA TNINLKHNTS 300
  TLNLDNIIIV IGNIKGNNNK DILNFKVHGT NLDNEMIIPA POKTHGTLNF KGNATLNGNI 360
  NNLNLIKFSG GHGKTLNLQG NTKVDNLVFA DSVLDSGTIS VNGLLDTDCV TFNNSNVNGG 420
  TLIINAKNTI SAKLLNATKA KIQINANLTM NHPSAGDISD IRIADNTIYT IDAKNGNVNL 480
65 LNNNAKIIIE GADSMALIN TGVTAERTFT IYNNLNQSGN DEYGIVKIEA IKKVITIANQ 540
  SGPYTIGQDN THRLKELIVE GAGDIIIDDT IFTKLLSINS TGQITFNRTL DLGAGGNIAF 600
  GKHGTLVVNG VTGSITTSEN NQGILTINSN NITGVIGTNE LGLKLVNIGA DPVTCANVF 660
```


| | | | | | | | |
|----|--|------------|-------------|------------|-------------|-------------|------|
| | ASVALTNPSS | VLILADGVTL | TGEVTTNNNT | KGVLSLGTGS | NITGQIGTNS | AALEKINIGA | 720 |
| | GASNIDSNIY | AGSTVLTDQT | SELTLNNDVV | VNSNIITTAG | NNSGKLIFTG | NGGITGNIGA | 780 |
| | NGAALQEVVF | NGTTNIGGTA | NSQNFTVAHS | AANVVITGLT | TGALKYKDTG | TIIAHGGLVG | 840 |
| | DIDFNNKAGK | FILGDGAMID | GSVLCNGGVA | GTLDFIGDGN | VTQNIADNA | NSISTINIQG | 900 |
| 5 | DNTKNVTIAN | DIFVDNIHFT | NGGILQLGGN | LTTHNIDFGA | NGGTLEFNGN | NTYNLNAIIV | 960 |
| | NGQNGILNAF | TNLKASDDTI | GTVKIINIGQ | IGTPQNFTIQ | VNNKNLTLVS | SVNSSINFGD | 1020 |
| | ANSQLILSAP | VDQTIKFINN | LNETGGGIIT | LDSNGNNLTI | SGNNGIKLGS | KGNELSSLNI | 1080 |
| | KGKVTVTNDL | DIQNIHQNLN | NNGALFDDQS | LTSAKIKNIN | IGTVAGGATY | TLDAINDNFD | 1140 |
| | LNTSGMVFKH | QDSILELKNS | SNTNDHTITL | TSALDPGNNQ | FGIIKLITDT | NKLTIDNNGN | 1200 |
| 10 | VAYTLGTANH | MLKQLTFASI | DNGAIALKVG | INVENVTLNI | KDIELNEVNA | NVLFNKNTTY | 1260 |
| | TATGNINGHV | DFQGNAGVIN | LNDDIEIDGS | VTSTGNVNGT | LNFNNGSGKVT | GLINNIVMLQ | 1320 |
| | AGAGDVSLSA | SGNYSITEIQ | GNGNNNLTF | ANSHLTDDIN | KTGGQDLNLV | FINGGSVSGS | 1380 |
| | IGANAAGVDI | IINAGSVNFS | NTLKSGNIVI | SDGATMQVNN | NVTATDISGK | NANNGTLKLN | 1440 |
| | NHTPINITST | LGNNNAIGTI | EVANNDVTIT | GTLOAQNIHF | SNATQAATLT | LGAASQVTNI | 1500 |
| 15 | TTAGNNIHTL | EVTDFDTGND | GIIGDANNRL | KSIELTGNGT | VTINSPIVYS | SITTANNAQG | 1560 |
| | NVKLNIEGGI | TYDLGSKIKS | LANVQISED | TIRGDVYSKY | LNIDAGKTIN | FDRGDNNMNP | 1620 |
| | KNLDIPDALI | DLVDLPRSL | LFNYFTDIKA | DNLNFADDTA | TANFKDAVVI | DAHIDNGGIL | 1680 |
| | KFNDNAWLTQ | EIKNANIEI | ASDKFMLLQK | NIKAATLIAD | NANLVLLDNV | EVNTNLNVRD | 1740 |
| | IVLDLANYEL | KYTGNVTHNG | LLTIITYFDT | ALQKGGHILV | SQGSNVDMSD | LDNLIKIKI | 1800 |
| 20 | HSDITNITSD | TKHQIVKLET | GAITYPVPQT | KVIIDASEEQ | NKFVKWVADA | NGLVLLTDTG | 1860 |
| | GRDDTGGRDD | TRGRGNTDNG | CRDNCVGN | SNNSSNEAGG | SSSDKNYGIT | DVVPIFDPS | 1920 |
| | ILDYTKNNYV | ASGIANQLIN | HVKDFGNTTD | AGKLLNDLGF | MSPNRVTETL | DRLSNRINVN | 1980 |
| | GLNEGVVGLN | GIEVENFLTD | IAINMDNFTA | KEIGNRLEEL | SDANTVNGLN | KTNTLLNNKI | 2040 |
| | NLKRLNTNNQ | AIIAAGDEDN | IVTGIWGMSE | YGKIKQNSKN | SASGYQSNTG | GGIIGFDYNI | 2100 |
| 25 | DNSIVIGAAY | TMADSKVKHK | NDKNGDRTKA | KSNIYSIYGL | YNWLTNNFFV | EAIGVYGRNK | 2160 |
| | IKNYEKRIIT | ITDQIAIGKF | INTFYSEYEL | GGYNYLISHR | TTITPMFGMR | YATFKNNGYK | 2220 |
| | ENNTTFQNL | IKKNYYDKFE | TILGLNSVTH | YLSQDIITKP | ELHWFINYQC | KNKLPNIDAR | 2280 |
| | LDGIDEPLTT | IRFKPAKITY | NLGGGISTKN | NMIEFGIRYN | LSLAKKYTAH | QGSCLKIKVNL | 2340 |
| 30 | <212> Type : PRT | | | | | | |
| | <211> Length : 2340 | | | | | | |
| | SequenceName : SEQ ID 171 | | | | | | |
| | SequenceDescription : | | | | | | |
| 35 | Sequence | | | | | | |
| | ----- | | | | | | |
| | <213> OrganismName : Rickettsia prowazekii | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | MAQKPNFLKK | IISAGLVTAS | TATIVAGEFSG | VAMGAAMQYN | RTTNAAATTF | DGIGFDQAAG | 60 |
| 40 | ANIPVAPNSV | ITANANNPIT | FNTPNGHLNS | LFLDTANDLA | VTINEDTTLG | FITNIAQQAK | 120 |
| | FFNFTVAAGK | ILNITGQGIT | VQEASNTINA | QNALTKVHGG | AAINANDLSG | LGSITFAAAP | 180 |
| | SVLEFNLINP | TTQEAPLTLG | ANSKIVNGGN | GTLNITNGFI | QVSDNTFAGI | KTINIDDCQG | 240 |
| | LMFNSTPDAA | NTLNLQVGGN | TINFNGIDGT | GKLVLVSKNG | AATEFNVGTG | LGGNLKGIIE | 300 |
| | LNTAAVAGKL | ISQGAANAV | IGTDNGAGRA | AGFIVSDNG | NAATISGQVY | AKNMVIQSAN | 360 |
| 45 | AGGQVTFEHI | VDVGLGGTTN | FKTADSKVII | TENSNGGSTN | FGNLDTQIVV | PDTKILKGNF | 420 |
| | IGDVKNNGNT | AGVITFNANG | ALVSASTDPN | IAVTNINAIE | AEGAGVVELS | GIHIAELRLG | 480 |
| | NGGSIFKLAD | GTVINGPVNQ | NALMNNNALA | AGSIQLDGSA | IITGDIGNGG | VNAALQHITL | 540 |
| | ANDASKILAL | DGANIIGANV | GGAIHFQANG | GTIKLTNTQN | NIVVNFDLDI | TTDKTGVVDA | 600 |
| | SSLTNNQTLT | INGSIGTVVA | NTKTLAQLNI | GSSKTILNAG | DVAINELVIE | NNGSVQLNHN | 660 |
| 50 | TYLITKTINA | ANQGQIIVAA | DPLNTNTTLA | DGTNLGSAEN | PLSTIHFATK | AANADSILNV | 720 |
| | GKGVNLYANN | ITTNDANVGS | LHFRSGGTSI | VSGTVGGQQG | HKLNNLILDN | GTTVKFLGDT | 780 |
| | TFNGGTKIEG | KSILQISNNY | TTDHVESADN | TGTLEFVNTD | PITVTLNKQG | AYFGVLKQVI | 840 |
| | ISGPGNIVFN | EIGNVGIVHG | IAANSISFEN | ASLGTSFLFP | SGTPLDVLTI | KSTVGNGTVD | 900 |
| | NFNAPIVVVS | GIDSMINNGQ | IIGDKKNIIA | LSLGSDNSIT | VNANTLYSGI | RTTKNNQGT | 960 |
| 55 | TLSGGMPNNP | GTIYGLGLEN | GSPKLKQVTF | TTDYNLNGSI | IANNVTINDY | VTLTTGGIAG | 1020 |
| | TDFDAKITLG | SVNGNANVRF | VDSTFSDPRS | MIVATQANKG | TVTYLGNALV | SNIGSLDTPV | 1080 |
| | ASVRFTGNDS | GAGLQGNISY | QNIDFGTYNL | TILNSNVILG | GGTTAINGEI | DLTNNLIFA | 1140 |
| | NGTSTWGDNT | SISTTLNVSS | GNIGQVVAIE | DAQVNATTTG | TTTIKIQDNA | NANFSGTQAY | 1200 |
| | TLIQGGARFN | GTLGAPNFAV | TGSNIFVKYE | LIRDSNQDYV | LTRTNDVLNV | VTTAVGNSAI | 1260 |
| 60 | ANAPGVSQNI | SRCLESTNTA | AYNNMLLAKD | PSDVATFVGA | IATDTSAAVT | TVNLNDTQKT | 1320 |
| | QDLLSNRLGT | LRYLSNAETS | DVAGSATGAV | SSGDEAEVSY | GVWAKPFYNI | AEQDKKGGIA | 1380 |
| | GKAKTTGVV | VGLDTLASDN | LMIGAAIGIT | KTDIKHQDYK | KGDKTDINGL | SFSLYGSQQL | 1440 |
| | VKNFFAQGNA | IFTLNKVKSK | SQRYFFESNG | KMSKQIAAGN | YDNMTFGGNL | IFGYDYNAMP | 1500 |
| | NVLVTPMAGL | SYLKSSNENY | KETGTTVANK | RINSKFSDRV | DLIVGAKVAG | STVNITDIVI | 1560 |
| 65 | YPEIHSFVVH | KVNGKLSNSQ | SMLDGQTAPF | ISQPDRTAKT | SYNIGLSANI | KSDAKMEYGI | 1620 |
| | GYDFNSASKY | TAHQGTLKVR | VNF | | | | 1643 |
| | <212> Type : PRT | | | | | | |

<211> Length : 1643
SequenceName : SEQ ID 172
SequenceDescription :

5 Sequence

<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :

| | | |
|----|--|------|
| 10 | MARIILEAHD VWEDGTGYQM LWDADHNQYG ASIPEESFWF ANG TIPAGLY DPFEYKVPVN | 60 |
| | ADASFSPTNF VLDGTASADI PAGTYDYVII NPNPGIIYIV GEGVSKGNDY VVEAGKTYHF | 120 |
| | TVQRQGPDA ASVVVTGEGG NEFAPVQNLQ WSVSGQTVTL TWQAPASDKR TYVLNESFDT | 180 |
| | QTLPNGWTMI DADGDGHNWL STINVYNTAT HTGDGAMFSK SWTASSGAKI DLSPDNYLVT | 240 |
| | PKFTVPENGK LSYWVSSQEP WTNEHYGVFL STTGNEAANF TIKLLEETLG SGKPAPMNLV | 300 |
| 15 | KSEGVKAPAP YQERTIDLSA YAGQQVYLAF RHFGCTGIFR LYLDDVAVSG EGSSNDYTYT | 360 |
| | VYRDNVVIAQ NLTATTFNQE NVAPGQNYNC VEVKYTAGVS PKVCKDVTVE GSNEFAPVQN | 420 |
| | LTGSAVGQKV TLKWDAPNGT PNPNGPTTTL SESFENGIPA SWKTIDADGD GNNWTTTPPP | 480 |
| | GGSSFAGHNS AICVSSASYI NFEGPQNPND YLVTPELSLP NGGTLTFWVC AQDANYASEH | 540 |
| | YAVYASSTGN DASNFANALL EEVLTAKTIV TAPEAIRGTR VQGTWYQKTQ QLPAGTKYVA | 600 |
| | FRHFGCTDFF WINLDDVEIK ANGKRADFTE TFESSTHGEA PAEWTTIDAD GDGQGWLCLS | 660 |
| 20 | SGQLGWLTAH GGTNVVASFS WNGMALNPDN YLISKDVTGA TKVKYKYAVN DGFPGDHYAV | 720 |
| | MISKGTGNAG DFTVVFEETP NGINKGGARF GLSTEANGAK PQSVWIERTV DLPAGTKYVA | 780 |
| | FRHYNCSLDN YILLDDIQFT MGSPTPTDY TYTVYRDGK IKEGLTETTF EEDGVATGNH | 840 |
| | EYCVVEKYTA GVSPKECVNV TVDPVQFNPV QNLTSAGVQ KVTWKWDAPN GTPNPNGTT | 900 |
| | TLSESFENGI PASWKTIDAD GDGNNWTTTP PPGGTSFAGH NSAICVSSAS YINFEGPQNP | 960 |
| 25 | DNYLVTPELS LPNGGTLTFW VCAQDANYAS EHYAVYASST GNDASNFAA LLEEVLTAKT | 1020 |
| | VVTAPEAIRG TRVQGTWYQK TVQLPAGTKY VAFRHFCTD FFWINLDDVE IKANGKRADF | 1080 |
| | TETFESSTHG EAPAEWTTID ADGDGQGWLC LSSGQLDWT AHGGTNVVAS FSWNGMALNP | 1140 |
| | DNYLISKDVT GATKVKYKYA VNDGFPGDHY AVMISKGTN AGDFTVVFEE TPNGINKGGA | 1200 |
| | RFGLSTEANG AKPQSVWIER TVDLPAGTKY VAFRHYNCSD LNYILLDDIQ FTMGGSPTPT | 1260 |
| 30 | DYTYTVYRDG TKIKEGLTET TFEEDGVATG NHEYCVVEKY TAGVSPKECV NVTVPVQFN | 1320 |
| | PVQNLTSAGV GQKVTWKWDA PNGTPNPNG TTTLSESFEN GIPASWKTID ADGDGNNWTT | 1380 |
| | TPPPGGTSFA GHNSAICVSS ASYINFEGPQ NPDNYLVTPE LSLPNGGTLT FWVCAQDANY | 1440 |
| | ASEHYAVYAS STGNDASNFA NALLEEVLTA KTVVTAPEAI RGTRVQGTWY QKTVQLPAGT | 1500 |
| | KYVAFRHFCT TDFFWINLDD VEIKANGKRA DFTETFESST HGEAPAEWTT IDADGDGQGW | 1560 |
| 35 | LCLSSGQLGW LTAHGGTNV ASFSWNGMAL NPDNYLISKD VTGATKVKY YAVNDGFPGD | 1620 |
| | HYAVMISKTG TNAGDFTVVF EETPNGINKG GARFGLSTEA NGAKPQSVWI ERTVDLPAGT | 1680 |
| | KYVAFRHYNC SDLNILLDD IQFTMGGSPT PTDYTYTVYR DGTKIKEGLT ETTFEEDGVA | 1740 |
| | TGNHEYCVVE KYTAGVSPKE CVNVTINPTQ FNPVQNLTA QAPNSMDAIL KWNAPASKRA | 1800 |
| | EVLNEDFENG IPASWKTIDA DGDGNNWTTT PPPGSSFAG HNSAICVSSA SYINFEGPQNP | 1860 |
| 40 | PDNYLVTPEL SLPGGGTLTF WVCAQDANYA SEHYAVYASS TGNDASNFAA ALLEEVLTAK | 1920 |
| | TVVTAPEAIR GTRVQGTWYQ KTVQLPAGTK YVAFRHFCT DFFWINLDDV VITSGNAPSY | 1980 |
| | TYTYRNNTQ IASGVTETTY RDPDLATGFY TYGVKVVYPN GESAIETATL NITSLADVTA | 2040 |
| | QKPYTLTVVG KTITVTCQGE AMIYDMNGRR LAAGRNTVVY TAQGGHYAVM VVVDGKSYVE | 2100 |
| | KLAVK | 2105 |

<212> Type : PRT
<211> Length : 2105
SequenceName : SEQ ID 173
SequenceDescription :

50 Sequence

<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :

| | | |
|----|---|-----|
| 55 | MKTSERILSY FFLCAVFSL GSCEGLYAQV TFPNYSPTAA SSIIVCSGEE TLIIDFTVVQ | 60 |
| | EDSNGIKVNV KLADGVEYVV GTAVSVTQG NAVTVAETNV SNPNEPVFTV KSADGNNVVE | 120 |
| | LGTIVKLTIK RRAVCTAWSN AINAAETGFV FKDKVTVTIG DHSDSKESNS YSVNYPNLTI | 180 |
| | KQPAPQVKNQ IGETIVREFS ITNGSQNPQT TVYLSIEYPD EAYLTGVGAM TLQAKLGASG | 240 |
| | TYADLTPTVT NGKVRIYTLS GSSLGPDHLL TNGEIIYLKE TFKLKTCAV TVYRVGWGCS | 300 |
| | IDSQCEIKTT AATITMAAGA ANITGYSVTG PDYRSPTFSL CQPFELTIKF SNSGAGGSMG | 360 |
| 60 | AAFNINTIGR NDYYRPRGFV LHEFIDVKVN GKPVTNFKTD GSELDLRFDG QFTEDPDGPG | 420 |
| | VGLDDVDGDG FYDDLPGVAT ITITVTVRLK CDQFTACNNA PNDLSRGLI LKTLYQTSCT | 480 |
| | RTSWIDPNTW FNLSSTHLYL SRESVQDASH MPTVIEKDTF FDLKIMTSY SILSSYNNIW | 540 |
| | YANPNTRYVV EIVFPQGMTM PPKSDIEWTN IKNHPIDGSL VFTPPINLPD ANITTSNTM | 600 |
| | TIVSPSQERG FVTLHGKVD CTNNHEMVVE YKIREVFNYL HFPDCLCPVG PIMCNTAKRY | 660 |
| 65 | VLGCDPPCGR GAETSVPKIE RADNSLGWTD YTMRTQSRN NISAYDLAKA LYMDEVNITA | 720 |
| | TSIQHGTASS LGARFVLATG VDRVETLTPL SADIKIFRDG VQIVSVGTYT TFRSIRNNN | 780 |
| | AEQVIDWDFI SILPAGGLLD RDKVDVVTRY RVTSQNAHRV DTQVGREWFY YNSTANVSPI | 840 |

WDEANPLTCL ILVPEIYIMG TFVVNGTDPH VISQCTPTDL GRVANHYARR FGSGAFEYAN 900
EYRPGVKIRN IYLKVPKSYT LNRVEYSNHR NHSSLGTTMP FEEINHTDVT SQGEYNIYKY 960
QLADNEKAHF NITVKNAIGA ALKVNVSPTC ASSAVATNYD KISYYVDYID YYYAATQPT 1020
VPNSLDIVAD QSAGSNGIYS VSALNVYNRP ILYTNKPSIA LVNQSGEVEL VGKTGEWKLR 1080
5 ISNPSSATAP YVWLALPTTS GLTIEKVTD AGETEMAFITY SGGKMYRLSE AGVPVGSALD 1140
YTIHFTYSGC SPIALKAMGG WNC SAYPLSL DEYVCSSQVI DLKLKPLPAA MELTEIAVPD 1200
PTAAATLCST LEYIYSIQST DNANVYSPTF SIFPEEGLVV TPNQVQVEYP AGSGNWAALN 1260
VVNNSVNLQ HPALTTIGYL KGLKEGESND NQRKILVKFY IKTECSFVSG KNFRVRADGR 1320
NACNQNAKGS GLAISTPPIR INGAIEPYTT SASTQLVTTT TSQSDCKAPK RVKVVQTVVG 1380
10 GETTPKAYLE ITLPLGFKYV TGSYAPDNTH PGGVNASPAG TEEVTLTANG EDKIKINVKA 1440
GLTSGQSFAY TLEMKEDDDN VPACGNHTIE IVNVEEIEGL WCEGVQCAET LVVTGANKFE 1500
FELDKPYLDI TVISAVSTFS GGKENLTIEY KVSNTSTTQP LKPGAVVTLF SDKDNNQVFS 1560
GGDVAVATQE LVAEITNTTP LTQIMKVKG V SSSHTGNLVL TILPKDGCYC EIKSPMVTLN 1620
HLPSNYWIGG TVGKPNEWKE PNNWTNDQVP DAAEDVEFAT EVNNPTDPNN PKSGPAKENL 1680
15 HLDDIHQNGT AGRVIGNLIN DSDKDLVITT GNQLTINGVV EDNNPNVGTI VVKSSKDNPT 1740
GTLLFANPGN NQNVGGTVEF YNQGYDCADC GMYRRSWQYF GIPVNESDFP YDHVDGNATV 1800
NQWVEPFNGD KWRPAPYAPD TKLQKFKGYQ ITNDVQAQPT GVVSFKGTLC VCD AFLNLTR 1860
TSGVNYSGAN LIGNSYTGA I DIKQGI VFPF EVEQTVYLFN TGTRDQWRKL NGSTVSGYRA 1920
GQYLSVPKNT AGQDNLPDRI PSMHSFLVKM QNGASCTLQI LYDKLLKNTT VNNGNGTQIT 1980
20 WRSNGSGSAN MPSLMDVLG NESADRLWIF TDGGLSFGFD NGWDGRKLTE KGLSQLYAMS 2040
DIGNDKFQVA GVP ELNLLI GFDADKDGQY TLEFALSDHF AKGGVFLEDL SRGVTRRVVD 2100
GGSYFDAKR GDSGARFRLS YDEEWVESAE VSVLVGTAGK RIVITNNSEH ACQANVYTTD 2160
GKLLIRLDVK PGSKSMTEPL VDG VYVSLQ SPATSSNVRK VVVN 2204

<212> Type : PRT
25 <211> Length : 2204
SequenceName : SEQ ID 174
SequenceDescription :

Sequence
30 -----

<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MNKFYKSLQ SGLAAVFSMA TALASAQIS FGGEPLSFSS RSAGTHSFDD AMTIRLTPDF 60
NPEDLIAQSR WQSQRDGRPV RIGQVIPVDV DFASKASHIS SIGDVDVYRL QFKLEGAKAI 120
35 TLYYDAFNIP EGGRLYIYTP DHEIVLGAYT NATHRRNGAF ATEPVPSEL IMDYEVSRGG 180
TLPDIKISGA GYIFDKVGR PVTDNHYGIG EDDSDSDCEI NINCEGADW QAEKNGVVQM 240
IMVKGQYISM CSGNLLNNTK GDFTPLIISA GHCSITTNF GVTQSELDKW IFTFHYEKRK 300
CSNGTLAIFR GNSIIGASMK AFLPIKGS D GLLQLNDEV PLRYRVYNG WDSTPDIPSS 360
GAGIHHFAGD AMKISILKKT PALNTWISSS GSGGTDDHFFY FKYDQGGTEG GSSGSSLFNQ 420
40 NKHVVGTLTG GAGNCGGTEF YGRLNSHWNE YASDGNSTRM DIYLDPQNG QTTILNGTYR 480
DGYKPLPSVP RLLQSTGDQ VELNWTAVPA DQYPSYQVE YHIFRNGKEI ATTKELSYSD 540
AIDESIIGSG IIRYEVSAF IYPSPLDGVE SYKDTDKTSA DLAIGDIQTK LKPDVTPPLG 600
GGVSLSWKVP FLSQLVSRFG ESPNPVFKTF EVPYVSAAAA QTPNPPVGVV IADKFMAGTY 660
PEKAAIAAVY VMPSAPDSTF HLFLKSNTNR RLQKVTTSPD WQAGTWLRIN LDKPFPVNND 720
45 HMLFAGIRMP NKYKLNRAIR YVRNPDNLF ITGKKISYNN GVSFEGYGIP SLLGYMAIKY 780
LVVNTDAPKI DMSLVQEPYA KGTNVAPFPE LVGIYVYKNG TFIGTQDPSV TTYSVSDGTE 840
SDEYEIKLVY KSGGISNGVA QIENNAVVA YPSVVTDRFS IKNAHMHVHAA ALYSLDGKQV 900
RSWNNLRNGV TFSVQGLTAG TYMLVMQTAN GPVSQKIVKQ 940
<212> Type : PRT
50 <211> Length : 940
SequenceName : SEQ ID 175
SequenceDescription :

Sequence
55 -----

<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MKNLNKFVSI ALCSSLLGGM AFAQQTTELGR NPNVRLLEST QQSVTKVQFR MDNLKFTEVQ 60
TPKGMAQVPT YTEGVNLSEK GMPTLPILSR SLAVSDTREM KVEVVSSKFI EKKNVLIAPS 120
60 KGMIMRNEDP KIPYVYGKS YSQNKFFPGE IATLDDPFIL RDVRGQVNF APLQYNPVTK 180
TLRIYTEITV AVSETSEQGK NILNKKGTFA GFEDTYKRMF MNYEPGRYTP VEEKQNGRMI 240
VIVAKKYEGD IKDFVDWKNQ RGLRTEVKVA EDIASPVTAN AIQQFVKQY EKEGNDLTYV 300
LLIGDHKDIP AKITPGIKSD QVYGQIVGND HYNEVFIGRF SCESKEDLKT QIDRTIHYER 360
NITTEDKWLQ QALCIASAE GSPADNGESD IQHENVIANL LTQYGYTKII KCYDPGVTPK 420
65 NIIDAFNGGI SLANYTGHGS ETAWGTSHFG TTHVKQLTNS NQLPFIFDVA CVNGDFLFSM 480
PCFAEALMRA QKD GKPTGTV AIIASTINQS WASPMRGQDE MNEILCEKHP NNIKRTFGGV 540
TMNGMFAMVE KYKKDGEKML DTWTVFGDPS LLVRTLVPTK MQVTAPAQIN LTDASVNVSC 600

| | | | | | | | |
|----|--|------------|------------|-------------|------------|-------------|------|
| | DYNGAIATIS | ANGKMFGSAV | VENGTATINL | TGLTNESTLT | LTVVGYNKET | VIKTINTNGE | 660 |
| | PNPYQPVSNL | TATTQGQKVT | LKWDAPSTKT | NATTNTARSV | DGIRELVLLS | VSDAPELLRS | 720 |
| | GQAEIVLEAH | DVWNDGSGYQ | ILLDADHDQY | GQVIPSDTHT | LWPNCSVPAN | LFAPFEYTVP | 780 |
| | ENADPSCSPT | NMIMDGTASV | NIPAGTYDFA | IAAPQANAKI | WIAGQGPTKE | DDYVFEAGKK | 840 |
| 5 | YHFLMKKMGs | GDGTELTISE | GGGSDYTYTV | YRDGTIKIEG | LTATTFEEDG | VAAGNHEYCV | 900 |
| | EVKYTAGVSP | KVCKDVTVEG | SNEFAPVQNL | TGSAVGQKVT | LKWDAPNGTP | NPNPNPNPNP | 960 |
| | NPGTTTLSES | FENGIPASWK | TIDADGDGHG | WKPGNAPGIA | GYNSNGCVYS | ESFGLGGIGV | 1020 |
| | LTPDNYLITP | ALDLPNGGKL | TFWVCAQDAN | YASEHYAVYA | SSTGNDASNF | TNALLEETIT | 1080 |
| | AKGVRSPETI | RGRIQGTWRQ | KTVDLPAGTK | YVAFRHFQST | DMFYIDLDEV | EIKANGKRAD | 1140 |
| 10 | FTETFESSTH | GEAPAEWTTI | DADGDGQGWL | CLSSGQLDWL | TAHGGTNVVS | SFSWNGMALN | 1200 |
| | PDNYLISKDV | TGATKVKYVY | AVNDGFPGDH | YAVMISKGTG | NAGDFTVVFE | ETPNGINKGG | 1260 |
| | ARFGLSTEAD | GAKPQSVWIE | RTVDLPAGTK | YVAFRHYNCS | DLNYILLDDI | QFTMGGSPTP | 1320 |
| | TDYTYTVYRD | GTKIKEGLTE | TTFEEDGVAT | GNHEYCVVEK | YTAGVSPKCC | VNVTVNSTQF | 1380 |
| | NPVKNLKAQP | DGGDVVLKWE | APSAKKTEGS | REVKRIGDGL | FVTIEPANDV | RANEAKVVLQ | 1440 |
| 15 | ADNVWGDNTG | YQFLLDADHN | TFGSVIPATG | PLFTGTASSD | LYSANFEYLI | PANADPVVTT | 1500 |
| | QNIIVTGQGE | VVIPGGVYDY | CITNPEPASG | KMWIAGDGGN | QPARYDDFTF | EAGKKYTFMT | 1560 |
| | RRAGMGDGTG | MEVEDDSPAS | YTYTVYRDGT | KIKEGLTETT | YRDAGMSAQS | HEYCVVEVKYT | 1620 |
| | AGVSPKVCVD | YIPDGVADVT | AQKPYTLTVV | GKTIITVTCQG | EAMIYDMNGR | RLAAGRNTVV | 1680 |
| | YTAQGGYYAV | MVVVDGKSYV | EKLAIK | | | | 1706 |
| 20 | <212> Type : PRT | | | | | | |
| | <211> Length : 1706 | | | | | | |
| | SequenceName : SEQ ID 176 | | | | | | |
| | SequenceDescription : | | | | | | |
| 25 | Sequence | | | | | | |
| | ----- | | | | | | |
| | <213> OrganismName : Porphyromonas gingivalis W83 | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | MKRKPLFSAL | VILSGFFGSV | HPASAQKVPA | PVDGERIIME | LSEADVECTI | KIEAEDGYAN | 60 |
| 30 | DIWADLNGNG | KYDSGERLDS | GEFRDVEFRQ | TKAIVYGKMA | KFLFRGSSAG | DYGATFIDIS | 120 |
| | NCTGLTAFDC | FANLLTELDL | SKANGLTFVN | CGKNQLTKLD | LPANADIETL | NCSKNKITSL | 180 |
| | NLSTYTKLKE | LYVGDNGLTA | LDLSANTLLE | ELVYSNNEVT | TINLSANTNL | KSLYCINNKM | 240 |
| | TGLDVAANKE | LKILHCNNNQ | LTALNLSANT | KLTTLSFFNN | ELTNIDLSDN | TALEWLFCNG | 300 |
| | NKLTKL DVSA | NANLIALQCS | NNQLTALDLS | KTPKLTTLNC | YSNRIKDTAM | RALIESLPTI | 360 |
| 35 | TEGEGRFVPY | NDDEGGEEEN | VCTTEHVEMA | KAKNWKVLTS | WGEPPPGITA | LISIEGESEY | 420 |
| | SVYAQDGILY | LSGMEQGLPV | QVYTVGGSM | YSSVASGSAM | EIQLPRGAAY | VVRIGSHAIK | 480 |
| | TAMP | | | | | | 484 |
| | <212> Type : PRT | | | | | | |
| | <211> Length : 484 | | | | | | |
| 40 | SequenceName : SEQ ID 177 | | | | | | |
| | SequenceDescription : | | | | | | |
| | Sequence | | | | | | |
| | ----- | | | | | | |
| 45 | <213> OrganismName : Shigella flexneri 2a str. 2457T | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | MKRAITLFAV | LLMGWSVNAW | SFACKTANGT | AIPIGGGSAN | VYVNLAPVVN | VGQNLVVDLS | 60 |
| | TQIFCHNDYP | ETITDYVTLQ | RGSAYGGVLS | NFSGTVKYSG | SSYPFPTTSE | TPRVVYNSRT | 120 |
| | DKPWPVALYL | TPVSSAGGVA | IKAGSLIAVL | ILRQTNNYNS | DDFQFVWNIY | ANNDVVVPTG | 180 |
| 50 | GCDVSARDVT | VTLPDYPGSV | PIPLTVYCAK | SQNLGYL LSG | TTADAGNSIF | TNTASFSPAQ | 240 |
| | GVGVQLTRNG | TIIPANNTVS | LGAVGTSAVS | LGLTANYART | GGQVTAGNVQ | SIIGVTFVYQ | 300 |
| | <212> Type : PRT | | | | | | |
| | <211> Length : 300 | | | | | | |
| 55 | SequenceName : SEQ ID 178 | | | | | | |
| | SequenceDescription : | | | | | | |
| | Sequence | | | | | | |
| | ----- | | | | | | |
| 60 | <213> OrganismName : Shigella flexneri 2a str. 2457T | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | MGIKQHNGNT | KADRLAELKI | RSPSIQLIKF | GAIGLNAIIF | SPLLIAADTG | SOYGTNITIN | 60 |
| | DGDRITGDTA | DPSGNLYGVM | TPAGNTPGNI | NLGNVDVTNV | NDASGYAKGI | IIQGNSSSLT | 120 |
| | ANRLTVDVVG | QTSAGINLI | GDYTHADLGT | GSTIKSNDDG | IIIGHSSTLT | ATQFTIENSN | 180 |
| 65 | GIGLTINDYG | TSVDLGSGSK | IKTDGSTGVY | IGGLNGNNAN | GAARFTATDL | TIDVQGY SAM | 240 |
| | GINVQKNSVV | DLGTNSTIKT | NGDNAHGLWS | FGQVSANALT | VDVTGAAANG | VEVRGGTTTI | 300 |
| | GADSHISSAQ | GGGLVTSSSD | ATINFSGTAA | QRNSIFSGGS | YGASAQTATA | VINMQNTDIT | 360 |

VDRNGSLALG LWALSGGRIT GDSLAITGAA GARGIYAMTN SQIDLTSDLV IDMSTPDQMA 420
IATQHDDGYA ASRINASGRM LINGSVLSKG GLINLDMHPG SVWTGSSLSLSD NVNGGKLDVA 480
MNSVWNVTS NSNLDTLALS HSTVDFASHG STAGTFTTLN VENLSGNSTF IMRADVVGEG 540
NGVNNRGDLL NISGSSAGNH VLAIRNQGSE ATTGNEVLTV VKTTDGAASF SASSQVELGG 600
5 YLYDVRKNGT NWELYASGTV PEPTPNPEPT PAPAQPPIVN PDPTPEPAPT PKPTTTADAG 660
GNYLNVGYLL NYVENRTLMO RMGDLRNQSK DGNIWLRSYG GSLDSFASGK LSGFDMGYSG 720
IQFGGDKRLS DVMPLYVGLY IDSTHASPDY SGGDGTARSD YMGMYASYMA QNGFYSDLVI 780
KASRQKNSFH VLDSQNNGVN ANGTANGMSI SLEAGQRFNL SPTGYGFYIE PQTQLTYSHQ 840
NEMAMKASNG LNIHLNHYES LLGRASMILG YDITAGNSQL NVYVKTGAIR EFSGDTEYLL 900
10 NDSREKYSFK GNGWNNGVGV SAQYNKQHTF YLEADYTQGN LFDQKQVNGG YRFSF 955

<212> Type : PRT
<211> Length : 955
SequenceName : SEQ ID 179
15 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
20 <400> PreSequenceString :
MSKFVKTAIA AAMVMGVFTS TATIAAGNNG TARFYGTIED SVCSIVPDDH KLEVDMGDIG 60
AEKLNNGTT TPKSFQIRLQ DCVFDTQETM TTTFTGTVSS ANSGNYTIF NTDTGAAFNN 120
VSLAIGDSLQ TSYKSGMGID QKIVKDTSTN KGKAKQTLNF NAWLVGAADA PDLGNFEANT 180
TFQITYL 187
25 <212> Type : PRT
<211> Length : 187
SequenceName : SEQ ID 180
SequenceDescription :

30 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKIKTLAIVV LSALSLSAA ALADTTTVNG GTIHFKEGVV NAACAVDAGS VDQTVQLGQV 60
35 RTASLKQAGA TSSAVGFNIQ LNDCTTVAT KAAVAFLGTA IDATRTDVLQ LQSSAAGSAT 120
NVGVQILDRT GNALTLDGAT FSAQTTLNNG TNTIPFQARY YAIGEATPGA ANADATFKVQ 180
YQ 182
<212> Type : PRT
<211> Length : 182
40 SequenceName : SEQ ID 181
SequenceDescription :

Sequence

45 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MASISSLGVG SGLDLSSILD SLTAAQKATL TPISNQSSSF TAKLSAYGTL KSALTTFQTA 60
NTALSKADLF SATSTTSSTT AFSATTAGNA IAGKYTISVT HLAQAQTLTT RTTRDDTKTA 120
IATSDSKLTI QQGDDKDPIT IDISAANSSL SGIRDAINNA KAGVSASIIN VGNGEYRLSV 180
50 TSNDTGLDNA MTLVSVGDDA LQSFMGYDAS ASSNGMEVSV AAQNAQLTVN NVAIENSSNT 240
ISDALENITL NLNDVTTGNQ TLTITQDTSK VQTAIKDWNV AYNLSLIDTFS SLTKYTAVDA 300
GADSQSSSNG ALLGDSTLRT IQTQLKSMLS NTVSSSSYKT LAQIGITTDG SDGKLELDAD 360
KLTAALKKDA SGVGALIVGD GKKTGITTTI GSNLTSWLST TGIKAATDG VSKTLNKLTK 420
DYNAASDRID AQVARYKEQF TQLDVLMTSL NSTSSYLTOQ FENNSNSK 468
55 <212> Type : PRT
<211> Length : 468
SequenceName : SEQ ID 182
SequenceDescription :

60 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MEGKADNVVL ENGGRLDVLV GHTATNTRVD DGGTLDVRNG GTATTVSMGN GGVLLADSGA 60
65 AVSGTRSDGK AFSIGGGQAD ALMLEKGSSF TLNAGDTATD TTVNGGLFTA RGGTLAGTTT 120
LNNGAILTSL GKTVNNDTLT IREGDALLQG GALTGNGSVE KSGSGTLTVS NTTLTQKAVN 180
LNEGTLTLND STVTTDVIAQ RGTALKLTGS TVLNGAIDPT NVTLASGATW NIPDNATVQS 240

VVDDLSHAGQ IHFTSTRITGK FVPATLKVKKN LNGQNGTISL RVRPDMAQNN ADRLVIDGGR 300
ATGKTIILNLV NAGNSASGLA TSGKGIQVVE AINGATTEEG AFIQGNKLQA GAFNYSLNRD 360
SDESWYLRSE NAYRAEVPLY ASMLTQAMDY DRILAGSRSH QTGVSGENNS VRLSIQGGHL 420
GHDNNGGIAR GATPESSGSY GFVRLEGDLI RTEVAGMSVT AGVYGAAGHS SVDVKDDDGGS 480
5 RAGTVRDDAG SLGGYLNLIH NASGLWADI V AQGTRHSMKA SSDNNDFRVR GWGWLGSLET 540
GLPFSITDNL MLEPQLQYTW QGLSLDDGQD NASYVKFGHG SAQHVRAGFR LGSHHDMNFG 600
KGTSSRDLR GSAKHSVREL PVNWWVQPSV IRTFSSRGDM SMGTAAAGSN MTFSPSQNGT 660
SLDLQAGLEA RVRENITLGV QASYAHSING SSAEGYNSQA TLNVTF 706
<212> Type : PRT
10 <211> Length : 706
SequenceName : SEQ ID 183
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MAFSQAVSGL NAAATNLDVI GNNIANSATY GFKSGTASFA DMFAGSKVGL GVKVAGITQD 60
FTDGTNTTNG RGLDVAISON GFFRLVDSNG SVFYSRNGQF KLDENRNLIN TQGLQLTGYP 120
20 VTGTPPTIQG GANPTNISIP NTLMAAKTTT TASMQINLNS SDPLPTVTPF SASNADSYNK 180
KGSVTVFDSQ GNAHDMSVYF VKTGDNNWQV YTDSSDPNS IAKTATTLEF NANGTLVDGA 240
MANNIATGAI NGAEPATFSL SFLNSMQQNT GANNIVATTQ NGYKPGDLVS YQINDDGTVV 300
GNNSNEQTQL LGQIVLANFA NNEGLASEGD NVWSATQSSG VALLGTAGTG NFGTLTNGAL 360
EASNVDLSKE LVNMIVAQRN YKSNAQTIKT QDQILNTRVN LR 402
25 <212> Type : PRT
<211> Length : 402
SequenceName : SEQ ID 184
SequenceDescription :

Sequence
30 -----
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKLVHMASGL AVAIALAACA DKSADIQTPA PAANTSISAT QQPAIQQPNV SGTWVIRQKV 60
35 ALPPDAVLTV TLSDAVLADA PSKVLAKAV RTEGKQSPFS FVLFPNPADV QPNARILLSA 120
AITVNDKLVF ITDTVQPVIN QGGTKADLTL VPVQQTAVPV QASGGATTTV PSTSPTQVNP 180
SSAVPAPTQY 190
<212> Type : PRT
<211> Length : 190
40 SequenceName : SEQ ID 185
SequenceDescription :

Sequence

45 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MIIKKS GGRW QLSLLASVVI SAFFLNTAYA WQOEYIVDTQ PGHSTERYTW DSDHQPDYND 60
ILSQRIQSSQ RALGLEVNLA EETPVDVTSS MSMGWNFPLY EQVTITGPVAA LHYDGTITSM 120
YNEFGDSTTT LTDPLWHASV SSLGWRVDSR LGDLRPWAIQI SYNQQFGENI WKAQSGLSRM 180
50 TATNQNGNWL DVTVGADMLL NQNIAYAAL TQAENTTNNS DYLYTMGVSA RF 232
<212> Type : PRT
<211> Length : 232
SequenceName : SEQ ID 186
55 SequenceDescription :

Sequence

60 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKWCKRGYVL AAMLALASAT IQAADVTITV NGKVVAKPCT VSTTNATVDL GDLYSFSLMS 60
AGAASAWHDV ALELTNCPVG TSRVTASFSG AADSTGYKYN QGTANIQLE LQDDSGNTLN 120
TGATKTVQVD DSSQSAHFPL QVRALTVNGG ATQGTIQA VI SITYTYS 167
<212> Type : PRT
65 <211> Length : 167
SequenceName : SEQ ID 187
SequenceDescription :

Sequence

5 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRAPLITGL LLISTSCAYA SSGGCGADST SGATNYSSVV DDVTVNQTDN VTGREFTSAT 60
LSSTNWQYAC SCSAGKAVKL VYMVSPVLT TGHQTGYKYL NDSLDIKTTL QANDIPGLTT 120
DQVSVNTRF TQIKSSTVYS AATQTGVCQG DTSRYGPVNI GANTTFTLYV TKPFLGSMTI 180
PKTDIAVIKG AWVDGMGSPS TGDFHDLVKL SIQGNLTAPQ SCKINQGDVI KVNFGFINGQ 240
10 KFTTRNAMPD GFTPVDVDIT YDCGDTSKIK NSLQMRIDGT TGVVDQYNLV ARRRSSDNVP 300
DVGIRIENLG GGVANIPFQN GILPVDPSGH GTVNMRAWPV NLVGGELETG KFQGTATITV 360
MVR 363
<212> Type : PRT
<211> Length : 363
15 SequenceName : SEQ ID 188
SequenceDescription :

Sequence

20 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MQKNAAH TYA ISSLLVLSLT GCAWIPSTPL VQGATSAQPV PGPTPVANGS IFQSAQPINY 60
GYQPLFEDRR PRNIGDTLTI VLQENVASAK SSSANASRDG KTNFGFDTVP RYLQGLFGNA 120
RADVEASGGN TFNGKGGANA SNTFSGTLTV TVDQVLVNGN LHVVG EKQIA INQGT EFIRF 180
25 SGVVNPRTIS GSNTVPSTQV ADARIEYVGN GYINEAQNMG WLQRFFLNLS PM 232
<212> Type : PRT
<211> Length : 232
SequenceName : SEQ ID 189
30 SequenceDescription :

Sequence

35 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRHLNLCYR LVWNHITGAF VVASELARAQ GKRGGVAVAL SLAAVTS LVPV LAADIVVHPG 60
ETVNGGTLVN HDNQFVSGTA DGVTVSTGLE LGPDS DENTG GQWIKAGGTG RNTT VTANGR 120
QIVQAGGTAS DTVIRDGGGQ SLNGLAVNTT LDNRGEQWVH GGGKAAGTII NQDGYQTIKH 180
GGLATGTIVN TGAEGGPESE NVSSGQMVGG TAESTTINKN GRQVIWSSGM ARDTLIYAGG 240
40 DQTVHGEAHN TRLEGGNQYV HNGGTATETL INRDGWQVIK EGGTAAHTTI NQKESCR 297
<212> Type : PRT
<211> Length : 297
SequenceName : SEQ ID 190
45 SequenceDescription :

Sequence

50 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MMMKTIKHLL CCAIAASALI STGVHAASWK DALSSAASEL GNQNSTTQEG GWSLASLTNL 60
LSSGNQALSA DNMNNAAGIL QYCAKQKLAS VTDAENIKNQ VLEKLGLNSE EQKEDTNYLD 120
GIQGLLTKTD GQQLNLDNIG TTPLAEKVKT KACDLVLKQG LNFIS 165
<212> Type : PRT
55 <211> Length : 165
SequenceName : SEQ ID 191
SequenceDescription :

Sequence

60 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MFKGQKTLAA LAVSLLFTAP VYAADEGSSE IHFKGEVIEA PCEIHQDDID KEVELGQVTT 60
SHINQSHSD AVAVDLLLVN CDLENSSNGS GKGISKVAVT FDSSAKTTGA DPILNNTSTG 120
65 EATGVGVRLM NKDQSNIVLG TATPDIDLAP TSSEQTLNFF AWMEQIDQAT PVT PGAVTAN 180
ATYVLDYK 188
<212> Type : PRT

<211> Length : 188
SequenceName : SEQ ID 192
SequenceDescription :

5 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MSAGSPKFTV RRIAALSLVS LWLAGCSDTS NPPAPVSSVN GNAPANTNSG MLITPPPKMG 60
10 TTSTAQQPQI QPVQQPQIQ TQQPQIQPVQ PVAQQPVQME NGRIVYNRQY GNIPKGSYSG 120
STYTVKKGDT LFYIAWITGN DFRDLAQRNN IQAPYALNVG QTLQVGNASG TPITGGNAIT 180
QADAAEQGVV IKPAQNSTVA VASQPTITYS ESSGEQSANK MLPNNKPTAT TVTAPVTVPT 240
ASTTEPTVSS TSTSTPISTW RWPTEGKVE TFGASEGGNK GIDIAGSKGQ AIIATADGRV 300
VYAGNALRGY GNLIIIKHND DYLSAYAHND TMLVREQQEV KAGQKIATMG STGTSSTRLH 360
15 FEIRYK GKSV NPLRYLPQR 379

<212> Type : PRT
<211> Length : 379
SequenceName : SEQ ID 193
SequenceDescription :

20

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
25 MIKFLSALIL LLVTTAAQAE RIRDLTSVQG VRQNSLIGYG LVVGLDGTGD QTTQTPFTTQ 60
TLNNMLSQLG ITVPTGTNMQ LKNVAAMVT ASLPPFGRQG QTIDVVVSSM GNAKSLRGGT 120
LLMTPLKGVD SQVYALAQGN ILVGGAGASA GSSSVQVNQL NGGRITNGAV IERELPSQFG 180
VGNTLNLQLN DEDFSMAQQI ADTINVRGY GSATALDART IQVRVPSGNS SQVRFLADIQ 240
NMQVNVTPQD AKVVINSRTG SVVMNREVTL DSCAVAQGNL SVTVNRQANV SQPDTPFGGG 300
30 QTVVTPQTQI DLRQSGGSLQ SVRSSASLNN VVRALNALGA TPMDLMSILQ SMQSAGCLRA 360
KLEII 365

<212> Type : PRT
<211> Length : 365
SequenceName : SEQ ID 194
SequenceDescription :

35

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
40 <400> PreSequenceString :
MKRSIIAAAV FSSFFMSAGV FAADVDTGTL TIKGNIAESP CKFEAGGDSV SINMPTVPTT 60
VFEGKAKYST YDDAVGVTS MLKISCPKEV AGVKLSLITN DKITGNDKAI ASSNDTVGDN 120
SDVLDVSAPF NIESYKTAEG QYAIPFKAKY LKLT DNSVQS GDVLSLVMR VAQD 174

<212> Type : PRT
<211> Length : 174
SequenceName : SEQ ID 195
SequenceDescription :

50 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MAVQKNVIKG ILAGTFALML SGCVTVPDAI KGSSTTPQQD LVRVMSAPQL YVGQEARFGG 60
55 KVVAVQNQQG KTRLEIATVP LD SGARPTLG EPSRGRIYAD VNGFLDPVDF RGQLVTVVGP 120
ITGAVDGKIG NTPYKFMVMQ VTGYKRWHLT QQVIMPPQPI DPWFYGGRGW PYGYGGGWY 180
NPGPARVQTV VTE 193

<212> Type : PRT
<211> Length : 193
SequenceName : SEQ ID 196
SequenceDescription :

60

Sequence

65 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MRNKPFYLLC AFLWLAVSRV LAADSTITIR GYVRDNGCSV AAESTNFTVD LMENAAKQFN 60

NIGATTPVVP FRILLSPCGN AVSAVKVGFT GVADSHNANL LALENTVSAA AGLGIQLLNE 120
QQNQIPLNAP SSAISWTTLT PGKPNTLNFY ARLMATQVPV TAGHINATAT FTLEYQ 176

5 <212> Type : PRT
<211> Length : 176
SequenceName : SEQ ID 197
SequenceDescription :

10 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKLTVAALA VTTLTSGSAF AHEAGEFFMR AGSATVRPTE GAGGTLGSLG GFSVTNNTQL 60
GLTFTYMATD NIGVELLAAT PFRHKIGTRA TGDIA TVHHL PPTLMAQWYF GDASSKFRPY 120
15 VGAGINYTTF FDNGFNDHGK EAGLSDSLK DSWGAGQVG VDYLINRDWL VNMSVWYMDI 180
DTTANYKLGG AQQHDSVRLD PWVFMFSAGY RF 212
<212> Type : PRT
<211> Length : 212
SequenceName : SEQ ID 198
20 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
25 <400> PreSequenceString :
MFFKRGKILS AGRLNKKS LG IVMFLSVGLL LAGCSGSKSS DTGTYSGSVY TVKRGDTLYR 60
ISRTTGTSVK ELARLNGISP PYTIEVGQKL KLGGAKSSSS TRKSTAKSTT KTASVTPSSA 120
VPKSSWPPVG QRCWLWPTTG KVIMPYSTAD GGNKGIDISA PRGTPIYAAG AGKVVYVGNQ 180
LRGYGNLIMI KHS EYITAY AHNDTMLVNN QOSVKAGQKI ATMGSTDAAS VRLHFQIRYR 240
30 ATAIDPLRYL PPQGS KPKC 259
<212> Type : PRT
<211> Length : 259
SequenceName : SEQ ID 199
SequenceDescription :

35 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
40 MAQVINTNSL SLITQNNINK NQSALSSSIE RLSSGLRINS AKDDAAGQAI ANRFTSNIKG 60
LTQAARNAND GISVAQTTEG ALSEINN NLQ RIRELTVQAS TGTNSDSLD SIQDEIKSRL 120
DEIDRVSGQT QFNGVNV LAK DGSMKIQVGA NDGQTITIDL KKIDSDTLGL NGFNVNGGGA 180
VANTAASKAD LVAANATVVG NKYTVSAGYD AAKASDLLAG VSDGDTVQAT INNGFGTAAS 240
ATNYKYDSAS KSYSFDTTTA SAADVQKYL T PGVGD TAKGT ITIDGSAQDV QISSDGKITA 300
45 SNGDKLYIDT TGR LTKNGSG ASLTEASLST LAANNTKATT IDIGGTSISF TGNSTTPDTI 360
TYSVTGAKVD QAAFDKAVST SGNNVDFTTA GYSVNGTTGA VTKGVDSVYV DNNEALTTSD 420
TVDFYLQDDG SVTNGSGKAV YKDADGKLTT DAETKAATTA DPLKALDEAI SSIDKFRSSL 480
GAVQNR LDSA VTNLNNTTTN LSEAQSRIQD ADYATEVSNM SKAQIIQQAG NSVLAKANQV 540
PQQVLSLLQG 550
50 <212> Type : PRT
<211> Length : 550
SequenceName : SEQ ID 200
SequenceDescription :

55 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
60 MKKIACLSAL AAVLAFTAGT SVAATSTVTG GYAQSDAQGQ MNKMGGFNLK YRYEEDNSPL 60
GVIGSFTYTE KSRTASSGDY NKNQYYGITA GPAYRINDWA SIYGVVGVGY GKFTTEYPT 120
YKHDTSDYGF SYGAGLQFNP MENVALDFS Y EQSRIRSVDV GTWIAGVGYR F 171
<212> Type : PRT
<211> Length : 171
65 SequenceName : SEQ ID 201
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
5 MKRNIIGGAF TLASLMLAGH ALAEDGVVNF VGEIVDTTCE VTSDTADQIV PLGKVSKNAF 60
SGVGSGLASPQ KFSIKLENC P ATYTQAAVRF DGTEAPGGDG DLKVGTPPLTA GNP GDFGTGTG 120
QAIAATGVGI RIFNQSDNSQ VKLYNDSAYT AIDAEGKAEM KFIARYVATN ATVTAGTANA 180
DSQFTVEYKK 190
<212> Type : PRT
10 <211> Length : 190
SequenceName : SEQ ID 202
SequenceDescription :

Sequence

15 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKSTLALVV MGIVASASVQ AAEIYNK DGN KLDVYGK VKA MHYMSDNASK DGDQSYIRFG 60
FKGETQINDQ LTGYGRWEAE FAGNKAESDT AQQKTRLAFA GLKYKDLGSF DYGRNLGALY 120
20 DVEAWTDMFP EFGGDSSAQT DNFMTKRASG LATYRNTDFF GVIDGLNLTL QYQGKNENRD 180
VKKQNGDGF G TSLTYDFGGS DFAISGAYTN SDRTNEQNLQ SRGTGKRAEA WATGLKYDAN 240
NIYLATFYSE TRKMTPTTGG FANKTQNF EA VAQYQDFDGL RPSLGYVLSK GKDIEGIGDE 300
DLVNYIDVGA TYYFNKNMSA FVDYKINQLD SDNKLININD DTVAVGMTYQ F 351
<212> Type : PRT
25 <211> Length : 351
SequenceName : SEQ ID 203
SequenceDescription :

Sequence

30 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MRKQWL GICI AAGMLA ACTS DDGQQQTVSV PQPAVCNGPI VEISGADPRF EPLNATANQD 60
35 YQRD GKS YKI VQDPSRFSQA GLAAIYDAEP GSNLTASGEA FDP TKLTA AH PTLPIPSYAR 120
ITNLANGRMI VVRINDRGPY GNDRVISLSR AAADRLNTSN NTKVRIDPII VAQDGSLSGP 180
GMACTTVAKQ TYALPAPDDL SGGAGTSSVS GPQGDILPVS NSTLKSEDPT GAPVTSSGFL 240
GAPTTLAPGV LEGSEPTPAP QPVVTASSTT PATSPAMVTP QAASQSASGN FMVQVGAVSD 300
QARAQQYQQQ LGQKFGVPGR VTQNGAVWRI QLGP FASKAE ASTLQQRLOT EAQLQSFITT 360
40 AQ 362
<212> Type : PRT
<211> Length : 362
SequenceName : SEQ ID 204
SequenceDescription :

Sequence

45 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
50 MKKKTIYQCV ILFFSL LNIH VGMAGPEQVS MHIYGNVVDQ GCDVATKSAL QNIHIGDFNI 60
SDFQAANTVS TAADLNIDIT GCAAGITGAD VLFSGEADTL APTLLKLTDT GSGSGMATGI 120
AVQILDAQSQ QEIPLNQVQP LTPLKAGDNT LKYQLRYKST KAGATGGNAT AVLYFDLVYQ 180
<212> Type : PRT
55 <211> Length : 180
SequenceName : SEQ ID 205
SequenceDescription :

Sequence

60 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKNKLLFMML TILGAPGIAA AAGYDLANSE YNFAVNELSK SSFNQA AII G QAGTNNSAQL 60
RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN DASISQ GAYG NTAMIIQKGS 120
65 GNKANITQYG TQKTAVVVQR QSQMAIRVTQ R 151
<212> Type : PRT
<211> Length : 151

SequenceName : SEQ ID 206
SequenceDescription :

Sequence

5 -----
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MMKFKKCLLP VAMLASFTLA GCQSNADDHA ADVYQTDQLN TKQETKTVNI ISILPAKVAV 60
DNSQNKRNAQ AFGALIGAVA GGVIGHNVGS GSNSGTTAGA VGGGAVGAAA GSMVNDKTLV 120
10 EGVSLTYKEG TKVYTSTQEG KECQFTTGLA VVITTTYNET RIQPNTKCPE KS 172

<212> Type : PRT
<211> Length : 172
SequenceName : SEQ ID 207
15 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
20 <400> PreSequenceString :
MQTKKNEIWV GIFLLAALLA ALFVCLKAAN VTSIRTESTY TLYATFDNIG GLKARSPVSI 60
GGVVVGRVAD ITLDPKTYLP RVTLEIEQRY NHIPDTSSLS IRTSGLLGEQ YLALNVGFED 120
PELGTAILKD GDTIQDTKSA MVLEDLIGQF LYGSKGDDNK NSGDAPAAAP GNNETTEPVG 180
TTK 183

25 <212> Type : PRT
<211> Length : 183
SequenceName : SEQ ID 208
SequenceDescription :

30 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
35 MAPLAFSAQS LAESLTVEQR LELLEKALRE TQSELKKYKD EEKKKYTPAT VNRSVSTNDQ 60
GYAANPFPTS SAAKPDVAVL KNEEKNASET GSIYSSMTLK DFSKFKVDEI GFSYNGYYRS 120
GWGTASHGSP KSWAIGSLGR FGNEYSGWFD LQLKQRYNE NGKRVDVAVM IDGNVGQQYS 180
TGWFGDNAGG ENFMQFSDMY VTTKGFLPFA PEADFWVGKH GAPKIEIQML DWKTQRTDAA 240
AGVGLENWKV GPGKIDIALV REDVDDYDRS LQNKQQINTH TIDLRYKDIP LWDKATLMVS 300
GRYVTANESA SEKDNQDNNG YYDWKDTWMF GTSLTQKFDK GGFNEFSFLV ANNSIARNFG 360
40 RYAGASPTT FNGRYYGDHT GGTAVRLTSQ GEAYIGDHFV VANAIVYSFG NNIYSYETGA 420
HSDFESIRAV VRPAYIWDQY NQTGVELGYF TQONKDANSN KFNESGYKTT LFHTFKVNTS 480
MLTSRLEIRF YATYIKALEN ELDGFTFEDN KDAQFAVGAQ AEIWW 525

<212> Type : PRT
<211> Length : 525
45 SequenceName : SEQ ID 209
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
50 MKKRILSAVL VSGVTLSSAT TLSAVKADDF DAQIASQDSK INNLTAAQQQA AQAQVNTIQG 60
QVSALQTQQA ELQAENQRLE AQSATLGOQI QTLSSKIVAR NESLKQQARS AQKSNAATSY 120
INAIINSKSV SDAINRVSAI REVVSANEKM LQQQEODKAA VEQKQENQA AINTVAANQE 180
55 TIAQNTNALN TQQAQLEAAQ LNLQAELTTA QDQKATLVAQ KAAAEAAARQ AAAAQAAAEA 240
KAAAEAKALQ EQAAQAQAAA NNNTQATDVS DQQAADANT QAAQTGDSTE QSAAQAVNNS 300
DQESTTATEA QPSASSASTA AVAANTSSAN TYPAGQCTWG VKSLAPWVGN YWNGGQWAA 360
SAAAAGYRVG STPSAGAVAV WNDGGYGHVA YVTGVQGGQI QVQEANYAGN QSIGNYRGWF 420
NPGSVSYIYP N 431

60 <212> Type : PRT
<211> Length : 431
SequenceName : SEQ ID 210
SequenceDescription :

65 Sequence

<213> OrganismName : Streptococcus mutans UA159

<400> PreSequenceString :
MKVKKTYGFR KSKISKTLG AVLGTVAAVS VAGQKVFAD TTTTSDVDTK VVGQTGTG NPA 60
TNLPEAQGSA SKEAEQSQNQ AGETNGSIPV EVPKTDLDQA AKDAKSAGVN VVQDADV NKG 120
TVKTAEEAVQ KETEIKEDYT KQAE DIKKT DQYKSDVA AH EA EVAKIKAK NQATKEQY EK 180
5 DMAAHKAEVE RINAANAASK TAYEAKLAQY QADLAAVQKT NAANQAAYQK ALAAYQAE LK 240
RVQEANAAAK AAYDTAVAA NAKNTEIAAA NEEIRKRNAT AKAEYETKLA QYQAE LKRVQ 300
EANAANEADY QAKLTAYQTE LARVQKANAD AKAA YEAAVA ANNAKNAALT AENTA IKQRN 360
ENAKATYEAA LKQYEADLAA VKKANAANEA DYQAKLTAYQ TELARVQKAN ADAKAA YEAA 420
VAANNAANAA LTAENTA IKK RNADAKADYE AKLAKYQADL AKYQKDLADY PVK LKAYEDE 480
10 QASIKAA LAE LEKHK NEDGN LTPESAQNLV YDLEPNANLS LT TDGKFLKA SAVDDAFSKS 540
TSKAKYDQKI LQLDDL DITN LEQSN DVASS MEL YGNFGDK AGWSTTVSNN SQVKG SVLL 600
ERGQSATATY TNLQNSYYNG KKISKIVYKY TVDPKSKFQG QKVWL GIFTD PTLGVFASAY 660
TGQVEKNTSI FIKNEFTFYD EDGKPINFDN ALLSVASLNR ENNSIEMAKD YTGK FVKISG 720
SSIGEKNGMI YATDTL NFRQ GQGGARWTMY TRASEPGSGW DSSDAPNSWY GAGAIRMSGP 780
15 NNSVTLGAIS STLVPADPT MAIETGKKPN IWYSLNGKIR AVNVPKVTKE KPTPPVKPTA 840
PTKPTYETEK PLKPAPVAPN YEKEPTPPTR TPDQAE PNKP TPPTYETEK LEPAPVEPSY 900
EAEPTPPTRT PDQAE PNKP PPTYETEKPL EPAPVEPSYE AEPTPPTPTP DQPEPNKPVE 960
PTYEVIPTPP TDPVYQDLPT PPSVPTVHFH YFKLAVQPQV NKEIRNNNDI NIDRTLVAQK 1020
SVVKFQLKTA DLPAGRDETT SFVLVDPLPS GYQFNPEATK AASPGFDVTY DNATNTVTFK 1080
20 ATAATLATFN ADLTKSVATI YPTVVGQVLN DGATYKNNFT LTVNDAYGIK SNVVRVTTPG 1140
KPNPDNPNPN NYIKPTKVNK NENGVIDGK TVLAGSTNYY ELTWDLDQYK NDRSSADTIQ 1200
KGFYYVDDYP EEALELRQDL VKITDANGNE VTGVSVDNYT NLEAAPQEIR DVL SKAGIRP 1260
KGAQIFRAD NPREFYDTYV KTGIDLKIVS PMVVKKQMGQ TGGSYENQAY QIDFGNGYAS 1320
NIIINNVPKI NPKKDVTLTL DPADTNNVDG QTIPLNTVFN YRLIGGIIPA DHSEELFEYN 1380
25 FYDDYDQTDG HYTGQYKVFA KVDITFKDGS IIKSGAELTQ YTTAEVD TAK GAITIKFKEA 1440
FLRSVSIDSA FQAESYIQMK RIAVGTFENT YINTVNGVTY SSNTVKT TTP EDPTDPTDPQ 1500
DPSSPRTSTV INYKPQSTAY QPSSVQETLP NTGVTNNAYM PLLGIIGLVT SFSLLGLKAK 1560
KD 1562
<212> Type : PRT
30 <211> Length : 1562
SequenceName : SEQ ID 211
SequenceDescription :

Sequence
35 -----
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MLTELKAVLK KPMLWITMVG VALVPALYNI IFLSSMWD PY GKVSDLPVAV VNKDKTATYE 60
GKKMTIGKDM TDNMVRNKSL DYHFVDSEKA QKGLEKGDY MIITLPEDLS QNAASVLTDE 120
40 PKKLTIPYQT SKGHSFVASK MSETAAKTLK ESVSKNITSS YTKSLFKNMS TLKTGLGSAA 180
NASQKIATGS KQLANGSQVM TDNLNLLSNS SQSFAQGTNT LYSGLTAYTG GVGQLSAGLN 240
NLNNGLTAYT NGVGQLANGS SQLSNQSQKL LGGVAQLANG SASIQQLVNA SSQLNQGLIK 300
LSTATGLSEE QVQQFSSLIN QLGTNLQSIQ NYSDNGTATT ANSPDLSTYL SAITTAQA I 360
VNSGNTSQQT TTNQSNALAA VQATGAYQRL SAEDQSEIAA ALANTGSSTT TTGADANAVS 420
45 QAQAILNNVQ SIQSALSTLQ TTTANTPTSP SASLTQIKNT ANSVLPSAAT SLTTLSSGLT 480
QAKTALDSQV VPVSTALANG TAQLGSTFST GANSLMTGVG QYTNAVDILN AGANTLAAKN 540
NQLTDGTSQL VNGANQLNSN SGQLTKGTAQ LANGANQIET GAGKLAAGGE SLTAGLTTLS 600
SGSGELSKAL STAKNKL SLV AVDNDNAKTL SSPVTIKHTD KDNVKTNGVG MAPYMMSAAL 660
MVMAISTNTI FRVALSGKQA KTLREWIDQK LAVNGLIAVT GAILLYFGVH IIGLSANFEL 720
50 KTLGLIILTS ITFMVLVTTL VTWHDKFGSF AALILLLLQL GSSAGTYPLA VTDKFFQVVN 780
PYLPMSYSVS GLRETISMAG TIGNQLLALS LFFLTFAALG LLIARRRIRS VKVA 834

<212> Type : PRT
<211> Length : 834
55 SequenceName : SEQ ID 212
SequenceDescription :

Sequence
60 -----
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MVSQKNKSKK GQSKTFTLIS NRINLLFFLI VALFTVLLLR LAQMQLYDAK FYKSKLTEST 60
TYTIKTSSPR GQIYDAKGVA LVENEVKEV AFTRSNMTA KDIKANAKKL ADMVTLTESK 120
VTKRQKDDY LADPKNYQKI VKKL PNKKY DNFGNNLTES KIYANAVKAV PNSAIDYSED 180
65 EKKIIHIFSQ MNATSVFNTA SLTTGDLTAE QIAVLATSKS DLKGISVKTD WERKTDKNSI 240
TSIIGKVSSQ KTGLPAEEAN NYVKKGYSLN DRVGTSYLEK QYENDLQGSR TVQAIKVNKE 300
GKIISDKTTA KGTKGKNLKL TLDLEFQKGV EQILNQYFNS ELASGNTKYS EGVYAVVLNP 360

NTGAVLSMAG LEHDLKTGEV SSNALGAVTE VFTPGSVVKG ATLTAGWENG VLSGNQVLND 420
QPIQFAGSSP INSWFTNGST PLTASQSLEY SSNTYMQVLA LKLMGQDYHS GMTLSTDGYK 480
EAMEKLRTAT AQYGLGVSTG IDLPGESKGY TPEHYDPSNV LTESFGQFDN YTAMQLAQYA 540
AAVANGGKRI APHLVEGIYD NNKTGGLGNL VQSIDTKVLN NVSISDDMG I IKEGFYNVV 600
5 NGGSYATGKT LAKGASVPIS AKTGTAEEAYV TGDDGKSVYT SNLNVVAYAP SSNPQIAVAV 660
VLPHETDLHG TTSHAITRDI INLYQKMYPM NQ 692
<212> Type : PRT
<211> Length : 692
SequenceName : SEQ ID 213
10 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
15 <400> PreSequenceString :
MTVLKYGLGI LLSAILLAI IGGLLFTYYV SSTPKLSEAK LKATNSSLVY DSNNNLIADL 60
GAEKRESISS DSIPMKLVNA VTSIEDHRFF KHRGVDIYRI IGAAWSNLLH KSTQGGSTLD 120
QQLIKLAYFS TKESDQTLKR KAQEVWLSLQ MEKKYTKEEI LTFYVNVKYM GNGNYGMRTA 180
AKSYYGKDLK DLSIAQLATL AGIPQAPTQY DPYAQPKAAT SRRNTVLSQM YKHKKITKRE 240
20 YDAAVATPIS DGLQELKRSS SYPKYMDNYL KQVISEVKKR TGQDIFSAGM KVYTNVNADA 300
QQYLWNIYNT DEYIAYPDDN FQVASTVMDV TNGKVIAQLG GRHQDTNVSF GTNQAVLTDR 360
DWGSTMKPIS AYGPALSEA FTTTAQMLND SVYYYPGTTT QVYDWDHRYN GWMTIQTAIQ 420
QSRNVPVRA IDAAGLD TAK GFLSGLGIDY PEMRYSNAIS SNTSSSEQKY GASSEKMAAA 480
YAAFSNGGTY YEPQYVKNIE FKDGTSETYD AKGNRAMKET TAYMMTDMK TVLTYGTGTE 540
25 AAIPGLYQAG KTGTSNYDDN ELVEMSEKLG INPYGLGTIA PDENFVGYP QYSMAVWTGY 600
KNRLMPVYGD SMKIAAQVYR TMMAYLSSSG NSDWTMPDGL YRSGGYLYLN GSSGSNSRYG 660
AAPATSSSSS SSSSSDSNNN DQNNNQTEA SSDSSSSSSD ATTSSNP 707
<212> Type : PRT
<211> Length : 707
30 SequenceName : SEQ ID 214
SequenceDescription :

Sequence

35 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKSKTAKITL LSSLALAAFG ATNVFADEAS TQLNSDTVAA PTADTQASEP AATEKEQSPV 60
VAVVESHTQG NTTTTTSQVT SKELEDAN ANQEGLEVTE TEAQKQPSVE AADADNKAQA 120
QTINTAVADY QKAKAEFPQK QEYQNKDFEK YQSDVKEYEA QKAAEQYKK EVAQGLASGR 180
40 VEKAQGLVFI NEPEAKLSIE GVNQYLTKA RQKHATEDIL QQYNTDNYTA SDFTQANPYD 240
PKEDTWFMMK VGDQISVTYD NIVNSKYNDK KISKVKINYT LNSSTNNEGS ALVNLFDHPT 300
KTIFIGAQTS NAGRNDKISV TMQIIFYDEN GNEIDLSGNN AIMSLSSLNH WTKYGDHVE 360
KVNLDGNEFV KIPGSSVDLH GNEIYSAKDN QYKANGATFN GDGADGWDAV NADGTPRAAT 420
AYYGAGAMTY KGEPTFTTVG GNDQNLPTTI WFATNSAVAV PKDPGAKPTP PEKPELKKPT 480
45 VTWHKNLVVE TKTEEVPPVT PPTTPDEPTP EKPKTPEDPQ SPVVAKSVSF RTARKGEMRV 540
RERDYQPTLP HAGAQQNGL ATLGAIATF AAATLIAARK KEN 583
<212> Type : PRT
<211> Length : 583
SequenceName : SEQ ID 215
50 SequenceDescription :

Sequence

55 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MEQKIFSKRK SKIAGLCGAI LTTTVVALAS GTVIEADETI EQPVAAETVS QADGDNPEQT 60
TSVQQETAPQ QTKTSQSSDA TVDSEESATS PSDEQTVSQN DSNSSSQIDQ TIADTNRSDS 120
DHISKTSAAAT TEDQEEKVNS AKAQTAAATN NQDTRYSAK AYGNSNFNKT LTFEGKNANV 180
ADVTYNGVRD EYIVVNDPSA PYVPNANEIA KYLKEYLTEL RNINNIAIPV PSVDQVMQKY 240
60 AQDRANEEAN EKNGLDHD TN LPIPNNLTWV AEDGHLDMDS SIQSKSQEGY TLASDKATAY 300
YLALNWFSY FNIYDDPNDG LKSFHGVSI LSDGGTGMGL GLASGQDNEK GMWYAQLEFG 360
GNDNEDNTND FSSLKNGKGE WVLYYKGPV KFLPNTTFWY VKKGTSPDAA STPHNSDKPS 420
FQSSKDLDPN FKADNRFQEG KEASVHQAIP ATFKSHRDEV GNKDQNSLSA QLPDTGVQKN 480
NQLALIALGT GLILLSGLLL SKRKSLK 507
65 <212> Type : PRT
<211> Length : 507
SequenceName : SEQ ID 216

SequenceDescription :

Sequence

5 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MTFEKQKHFS LRKLKFG LVS VAIIAFLFAV TKTAEAD ETV ITEQRQTSKI NASSQKVENQ 60
TSNQVEAKTD SANKDPQ EKT GSVATDAPSM NSANNMSQSD KQNTVNEISS DSQQT KTDEQ 120
TDL PQNSFKQ QSAHV KMTTE AEKTPSHSIN T FVNDGNGNW Y YLGADGRNV TGSHTIGGKT 180
10 MYFAQDGKQV KGAF AQDSDG NKHY YDRDSG EMWTNR FVND QGNWYYLNND GVPVTGSITV 240
NGQSLYFNSD GSQVKGNFVE EDGSL RYYDK NSGDLLRKTS RTINGVNYQF DNDGNARAID 300
KIEVVKTSLV VDSYEFGPSV SKIILEFNHK VTPAVVHAGA MVTTAGVQRK ILNSYVSNAS 360
GHVVYFDSSH YVTELDIPY DPNDSSRNAS PFIFDSAAFR NNWVNSYTVK VDNLQVQADG 420
SNSSQIISSE QDAINN RFLP TTD RFSERGS YGNFN YAA YQ PEAAIGGEKN PLIVWLHGIG 480
15 EVGTDINIP L LASNVARLTE DPIQSHFTST GSGGQKGAYV LVPQSSIPWS QNQTASLMAL 540
IKAYVASHPD IDSRR IYLAG VSNGGGMTLD MGVA YPNYFA ALVPIAASYS NQLTDNQITA 600
AALKALKGQP MWLIHTRTDK TISADSSVLP FYKELLQAGA QNKWLSYYET NVGKHHS GVT 660
YNGHWSWIYF LNDQVTGTQN TDNAKNWSGL SGMVATNPTY GGD AKATVNG RTYSNVFDWL 720
NGQRRR 726
20 <212> Type : PRT
<211> Length : 726
SequenceName : SEQ ID 217
SequenceDescription :

25 Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKIFIKKHQQ SILYYSLSFL LPSFIMFLVL FSKNIYW GSS TTILASDGFH QYVIFDALFR 60
30 NILHGTDSL F YSFKAGLGFN IFALTSYYLG SFLT PFTYFF NVKNMADAFY LFTLIKFG LI 120
GLSAFYSLGQ IYTKISKSLV LMLSTSYALM SFTSSQLELN NWLDV FILLP LIMLGLQRLV 180
EKRGIFLYFL TLTCLFIQNY YFGFMTAIFL TLWFFTQVSW DIRNRMKRLS DFVLVSIFAT 240
LTSAFMLLPT FLDLKSHGEV LTEQISLFSS DIWYFDFFAK SLLGSYD TTK YGSIPTIYIG 300
LLPLIFAIF FVKS IWKQV KVAYFLLLAI IIASFIFQPL DLFWQGMHSP NMFLHRYSWA 360
35 FSLVIVIMAA ETLTRIKDIK LKNFYPAFTF LGVGLLATFL FKDYNNYLTQ VNFILTTIFL 420
VSYFIILFTF FNQLVSYKVI ISFTLIFTSF EIALNTFYQI EGIQTDWNFP SREVYEDNVK 480
EIDNYVKKTK KDNLEFFRTE KQIPQTYNDG MKFNYSNISQ FSSVKNNLSA QLLNSLGYYS 540
QGNHSTISYP NNTILMDSL F SIKYNINNQN PHKFGFHLKQ KNNKLQLYKN FYSLPLALMS 600
NHIYKDV KFD SYPLDNQ QKF VNELTDLNLT LFKEIPIISS VGMQVLDNRV TINGSKGNKA 660
40 QVYYTVKCPA NSQLYISLPN LTVNNK DENV FITTNKHTSS YIIDESYYLF NLGNYKKTQT 720
LIFKLSFPKN KTVSYDLPHI YALDLTAYQK SIKQLKSQTV KTTTKKNKIF TTYVAKKRTS 780
LIYTLPYDKG WFAKQNGKAI KISKAQNGLM KIDVSKGSGK IIMTFVPQGL YQGILLTCLG 840
IFLFV FYQLY YKKFNLK 857
<212> Type : PRT
45 <211> Length : 857
SequenceName : SEQ ID 218
SequenceDescription :

Sequence

50 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKLKHILRIG AVAFASILLL TACGSKTSKK TVTLATVGTT NPFSYEKKKGK LTGYDIEVAK 60
EVFKASDKYD VKYQKTEWTS IFSGLDSDKY QIGANNISYT KERANKYLYS NPTASNPLVL 120
55 VVPKDS DIKS YNDIAGHSTQ VVQGNTTVSM LQKF NKNHEN NQVKLNFTSE DLAHQIRNVS 180
DGKYDFKIFE KISAETI IKE QGLDNLKVID LPSDQKP YVY FIF AQDQKDL QKFVNKRLKK 240
LYENG TLEKL SKKYLGGSYL PDKKDMK 267
<212> Type : PRT
<211> Length : 267
60 SequenceName : SEQ ID 219
SequenceDescription :

Sequence

65 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MRFLVFLIAF FAFYKFIET ERIDSNTVAV NPDSLILKRF LKTNQLNGIM IVTGPDGKAQ 60

VFSNQSKVDG SPVSIKDYFP LASLQKLITG VAIQQQLIDKG KLSLNTPLSK YYPQIENSEN 120
ITIQNLLTHT SGLADRKEVP QQVLTQTQEQQ LDFSLTNYRV TYRKKWKYAN INYALLAGII 180
SQISGQNYAT YVRQHFLTAG KGWHFKKYIQ IKDKSKLAAL SVMDOSTTWD KLSKEVTSTF 240
GAGDYASRPV DYWKFMMAFI NDQFVPVSEY QRSMKMTSKS YYGGLYISQK MLHANGGGFD 300
5 TYSCFAYSNP KTKQVMVLF I TNGKYKRVKS LAAKAFKLYA DSYALRKNET SK 352

<212> Type : PRT
<211> Length : 352
SequenceName : SEQ ID 220
10 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
15 <400> PreSequenceString :
MKKKIALAAL SFVSAAVLAA CSSAPGGSSD AAGNKIGDTV KIGYNLELSG DVAAYGQAEK 60
NGANLAVEEI NKAGGIDGKK IKVISKDNKS DNGEASTIST NLATQSKVNA ILGPATSGAT 120
AAAAPNANDA AVPLVTPSGT QDNLTYSKKG VQDYIFRTTF QDSFQKIIA KYATDNLKAK 180
KVALYYDKSS DYAQGIADAF KKAYKGKITV EDTFQAKDQD FQAALTKFKN KDFDAIVIPG 240
20 YYTETGLITK QARDMGLTQP ILGPDGFNDE KYVEGAGAAN TNNVHYVSGY STKVALTNKA 300
EKFLKDYKAK YGEEPNMFAA LAYDSVYMIA DAAKDAKTSK DIATNLAKLK NFKGVTGKMT 360
IDKKHNPVKS AVMVGLKDGK EDTATAVEAK 390
<212> Type : PRT
<211> Length : 390
25 SequenceName : SEQ ID 221
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
30 <400> PreSequenceString :
MKKLSLLLLV CLSLLGLFAC TSKKTADKKL TVVATNSIIA DITKNIAGNK VVLHSIVPVG 60
RDPHEYEPPL EDVKKTSQAD VIFYNGINLE NGGNAWFTKL VKNAHKKTDK DYFAVSDSVK 120
TIYLENAKEK GKEDPHAWLD LKNGIYYAKN IMKRLSEKDP KNKSYQKNF QAYSAKLEKL 180
35 HKVAKEKISR IPTEKKMIVT SEGCFKYFSK AYDIPSAYIW EINTEEGTP NQIKALVKKL 240
RKSRVSALFV ESSVDDRPMK TVSKDTGIPI AAKIFTDSVA KKGQAGDSYY AMMKWNIDKI 300
ANGLSQ 306
<212> Type : PRT
<211> Length : 306
40 SequenceName : SEQ ID 222
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
45 <400> PreSequenceString :
MFVHTKTKKK RKWQRKVFL LLLFLLPIVS VLA FIVLFIG GGTAESH DVE ATTGGVKLSA 60
KQFADKTKLG ISEEEAKNAL AFADRLMSRH HFTAQATAGV LAVGFRESGF DVKAVNNSGG 120
VAGFFQWSGW GSSVNGDRWK VASKRELTE VEVDLMSTEL DGRYADVVK VGSATDEKQA 180
50 AKDWSQYYEG VAVSDGQTKA DKIESWATTI CEALKSGGTN YAKVNNTGTS STAIPQGWEN 240
ISAFDGHAYE GSENYPPGQC TWYVYNRAKQ LGVSFSPYMG NGGQWYQVQG YHSSHTPKAH 300
TALS FVNGQA GSDPTYGHVA FVEAVKDDGS ILISEMNVYG QPAMTVAYRT FDAETAKQFW 360
YVEGK 365
<212> Type : PRT
55 <211> Length : 365
SequenceName : SEQ ID 223
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
60 <400> PreSequenceString :
MKMKRKL LSL VSVLTILLGA FWVT KIVKAD QVTNYTNTAS ITKSDGTALS NDPSKAVNYW 60
EPLSFSNSIT FPDEVSIKAG DTLTIKLPEQ LQFTTALTFD VMHTNGQLAG KATDPNTGE 120
65 VTVTFTDIFE KLPNDKAMTL NFNAQLNHN ISIPGVVNFN YNNVAYSSYV KDKDITPISP 180
DVNVKGYQDK SNPGLIHWKV LINNKQGAID NLTLTDVUGE DQEIVKDSL AARLQYIAGD 240
DVDSLDEAAS RPYAEDFSKN VTYQTNDLGL TTGFTYTIPG SSNNAIFISY TTRLTSSQSA 300

GKDVSNTIAI SGNNINYSNQ TGYARIESAY GRASSRVKRQ AETTTVTETT TSSSSETTTS 360
EATTETSSTT NNNSTTTETA TSTTGASTTQ TKTTASQTNV PTTTNITTTTS KQVTKQKAKF 420
VLPSTGEQAG LLLTTVGLVI VAVAGVYFYR TRR 453
<212> Type : PRT
5 <211> Length : 453
SequenceName : SEQ ID 224
SequenceDescription :

Sequence
10 -----
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MTFKKLVLGL LSFVAVFTLV ACSSSNSKNL QDDIKEKKKL VVAVSPDYAP FEFKALVNGK 60
DTVVGADIDL AKATAKELGV KLELSSMSFD NVLSSLKTGK ADIAISGLSY TKERAQAYDF 120
15 SEAYYKTENA ILIKKSDLNK YTMISSEFNNK TKVAVQKGTI EEGLAKNQLK QSNITSLTSM 180
GEAVNELKSG QVDAIDLEKP VAEGYVSQNS DLVLAKVALK TGEDAKAVA LPKDSGQLVK 240
TVNKVIKKLK KEDKYQFIS DAVKLTGQQV D 271
<212> Type : PRT
<211> Length : 271
20 SequenceName : SEQ ID 225
SequenceDescription :

Sequence

25 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKHFFMTFS LLLAAVFLVA CSNLSDSGQR NWDKINKRGM LKIATAGTLY PQSYHDDHKN 60
LTGYDVEILK EIGKRLGLKV QFTEMGVDGM LTAIKSGQID VANYSLEDGN KNISKFLRTS 120
PYKYSFTSMV VRSKDDSGIH SWSDLKGKKA AGAASTNYMK IAKKLGAJLV VYDNVTNDVY 180
30 MKDLVNGRTD VIINDYYLQK IAAVAVKDKY AIKINQGLYA NPYSTSFTLS LKNKVLQKKI 240
NKAVKDMRKD GTLTKLSKKF FQGEDVTKKH YNSYKKIDIS DVD 283
<212> Type : PRT
<211> Length : 283
SequenceName : SEQ ID 226
35 SequenceDescription :

Sequence

40 <213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKLLKKMMQV ALATFFFGLL GTSTVFADDS EGWQFVQENG RTYYKKGALK ETYWRVIDGK 60
YYYFDPLSGE MVVGWQYIPA PHKGVTIGPS PRIEIALRPD WFYFGQDGV L QEFVGKQVLE 120
AKTATNTNKH HGEEYDSQAE KRVYYFEDQR SYHTLKTGWI YEEGYWYYLQ KDGGFDSRIN 180
RLTVGELARG WVKDYPLTYD EEKKAAPWY YLDPATGWQN LGNKWYYLRS SGAMATGWYQ 240
45 EGSTWYYLNA SNGDMKTGWF QVNGNWWYAY DSGALAVNTT VGGYYLNYNG EWVK 294
<212> Type : PRT
<211> Length : 294
SequenceName : SEQ ID 227
50 SequenceDescription :

Sequence

55 <213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKLLKKMMQV LLAVFFFGLL ATNTVFANTT GGRFVDKDNK KYVVKDDHKA IYWHKIDGKT 60
YYFGDIGEMV VGWQYLEIPG TGYRDNLFND QPVNEIGLQE KWYYFGQDGA LLEQTDKQVL 120
EAKTSENTGK VYGEQYPLSA EKRTYYFDNN YAVKTGWIYE DGNWYYLNKL GNF'GDDSYNP 180
LPIGEVAKGW TQDFHVTIDI DRSKPAPWYY LDASGKMLTD WQKVNGKWYY FGS'GSMATG 240
60 WKYVRGKWYY LDNKN'GDMKT GWQYLG'NKWY YLRSSGAMVT GWYQDGLTWY YLNAGNGDMK 300
TGWFQVNGKW YYAYSSGALA VNT'TVDGYSV NYNGE'WVQ 338
<212> Type : PRT
<211> Length : 338
SequenceName : SEQ ID 228
65 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
5 MNKKKMILTS LASVAILGAG FVASQPTVVR AEESPVASQS KAEKDYDAAK KDAKNAKKAV 60
EDAQKALDDA KAAQKKYDED QKKTEEKAAL EKAASEEMDK AVAAVQQAYL AYQQATDKAA 120
KDAADKMIDE AKKREEEAKT KFNTVVRAMVV PEPEQLAETK KKSEEAKQKA PELTKKLEEA 180
KAKLEEAEKK ATEAKQKVDA EEVAPQAKIA ELENQVHRLE QELKEIDese SEDYAKEGFR 240
APLQSKLDAK KAKLSKLEEL SDKIDELDAE IAKLEDQLKA AEENNNVEDY FKEGLEKTIA 300
AKKAELEKTE ADLKKAVNEP EKPAPAPETP APEAPAEQPK PAPAPQPAPA PKPEKPAPQP 360
10 KPEKTDDQQA EEDYARRSEE EYNRLTQQQP PKAEKPAPAP KTGWKQENGM WYFYNTDGSM 420
ATGWLQNNGS WYYLNSNGAM ATGWLQYNGS WYVLNANGAM ATGWAKVNGS WYVLNANGAM 480
ATGWLQYNGS WYVLNANGAM ATGWAKVNGS WYVLNANGAM ATGWLQYNGS WYVLNANGAM 540
ATGWAKVNGS WYVLNANGAM ATGWVKDGD T WYVLEASGAM KASQWFKVSD KWYVYVNGLGA 600
LAVNTTVDGY KVNANGEWV 619
15 <212> Type : PRT
<211> Length : 619
SequenceName : SEQ ID 229
SequenceDescription :
20 Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
25 MKILPFIARG TSYYLKMSVK KLVPLVVG LMLAAGDSVYA YSRGNGSIAR GDDYPAYYKN 60
GSQEIDQWRM YSRQCTSFVA FRLSNVNGFE IPAAYGNANE WGHRRAREGY RVDNTPPTIGS 120
ITWSTAGTYG HVAWVSNVMG DQIEIEEYNY GYTESYNKRV IKANTMTGFI HFKDLDSGSV 180
GNSQSSASTG GTHYFKTKSA IKTEPLVSAT VIDYYPGEK VHYDQILEKD GYKWLSTAY 240
NGSYRYVQLE AVNKNPLGNS VLSSTGGTHY FKIKSIAKTE PLVSATVIDY YYPGEKVHYD 300
QILEKDGKW LSYTAYNGSR RYIQLEGVTS SQNYQNQSGN ISSYGSNNSS TVGWKKINGS 360
30 WYHFKSNGSK STGWLKDGSS WYVLKLSGEM QTGWLKENG S WYVLGSSGAM KTGWYQVSGE 420
WYYSYSSGAL AINTTVDGYR VNSDGERV 448
<212> Type : PRT
<211> Length : 448
SequenceName : SEQ ID 230
35 SequenceDescription :
Sequence

<213> OrganismName : Streptococcus pneumoniae R6
40 <400> PreSequenceString :
MFASKSERKV HYSIRKFSIG VASVAVASLV MGSVVHATEN EGSTQAATSS NMAKTEHRKA 60
AKQVVDEYIE KMLREIQ LDR RKHTQONVALN IKLSAIKTKY LRELNVLEEK SKDELPSEIK 120
AKLDAAFEKF KKD TLKPGEK VAEAKKKVEE AKKKAEDQKE EDRRNYPTNT YKTLELEIAE 180
FDVKVKEAEL ELVKEEAKES RNEG TIKQAK EKVESKKA EA TRLENIKTDR KKAEEEEAKRK 240
45 ADAKLKEANV ATSDQGKPKG RAKRGVPGEL ATPDKKENDA KSSDSSVGEE TLPSSSLKSG 300
KKVAEAEKKV EEAEKKAKDQ KEEDRRNYPT NTYKTL DLEI AESDVKVKEA ELELVKEEAK 360
EPRDEEKIKQ AKAKVESKKA EATRLENIKT DRKKAEEEEAK RKA AEEDKV K EKPAPQPAPA 420
PATQPEKPAP KPEKPAPQPK AEKTDQQA E EDYARRSEEE YNRLTQQQPP KTEKPAQPST 480
PKTGWKQENG MWYFYNTDGS MATGWLQNNG SWYVLNANGA MATGWLQNNG SWYVLNANGS 540
50 MATGWLQNNG SWYVLNANGA MATGWLQYNG SWYVLNSNGA MATGWLQYNG SWYVLNANGD 600
MATGWLQNNG SWYVLNANGD MATGWLQYNG SWYVLNANGD MATGWVKDGD TWYVLEASGA 660
MKASQWFKVS DKWYVYVNGSG ALAVNTTVDG YGVNANGEWV N 701
<212> Type : PRT
<211> Length : 701
55 SequenceName : SEQ ID 231
SequenceDescription :
Sequence

60 <213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKKTILSLT TAAVILAAYV PNEPILAAV PNEPILADTP SSEVIKETKV GSIIQQNNIK 60
YKVLTVEGNI GTVQVNGVT PVEFEAGQDG KPFTIPTKIT VGDKVFTVTE VASQAFSYYP 120
DETGRIVYYP SSITIPSSIK KIQKKGFHGS KAKTIIFDKG SQLEKIEDRA FDFSELEEIE 180
65 LPASLEYIGT SAFSFSQKLK KLTFSSSSKL ELISHEAFAN LSNLEKLTLP KSVKTLGSNL 240
FRLTSLKHV DVEEGNESFA SVDGVLFSKD KTQLIYYP SQ KNDES YKTPK ETKELASYSF 300
NKNSYLKKE LNEGLEKIGT FAFADAIKLE EISLPNSLET IERLAFYGNL ELKELILPDN 360

VKNFGKHVMN GLPKFLTLTG NNINSLPSFF LSGVLDLKE IHIKNKSTEF SVKKDTFAIP 420
ETVKFYVTSE HIKDVLKSNL STSNDIIVEK VDNKQETDV AKPKKNSNOG VVGWVKDKGL 480
WYYLNEGSMS ATGWVKDKGL WYYLNEGSMS ATGWVKDKGL WYYLNEGSMS ATGWVKDKGL 540
WYYLNEGSMS ATGWVKDKGL WYYLNEGSMS ATGWVKDKGL WYYLNEGSMS ATGWVKDKGL 600
5 WYYLNEGSMS ATGWVKDKGL WYYLNEGSMS ATGWVKDKGL WYYLNEGSMS ATGWVKVSGK 660
WYYTYNSGDL LVNTTTPDGY RVNANGWVG 690
<212> Type : PRT
<211> Length : 690
SequenceName : SEQ ID 232
10 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pneumoniae R6
15 <400> PreSequenceString :
MEINVSRLRT DLPQVGVQPY RQVHAHSTGN PHSTVQNEAD YHWRKDPELG FFSHIVGNGC 60
IMQVGPVDNG AWDVGGGUNA ETAAVELIE SHSTKEEFMT DYRLYIELLR NLADEAGLPK 120
TLDTGSLAGI KTHEYCTNNQ PNNHSDHVDY PYLAKWGIS REQFKHDIEG GLTIETGWQK 180
NDTGYYVYHS DGSYPKDKFE KINGTWYYFD SSGYMLADRW RKHTDGNWYW FDNSGEMATG 240
20 WKKIADKWY FNEEGAMKTG WVKYKDTWY LDKAGAMVS NAFIQSADGT GWYYLKPDTG 300
LADRPEFTVE PDGLITVK 318
<212> Type : PRT
<211> Length : 318
SequenceName : SEQ ID 233
25 SequenceDescription :

Sequence

<213> OrganismName : Neisseria meningitidis Z2491
30 <400> PreSequenceString :
MTFAYWCILI AYLLPLFCAA YAKKAGGFRF KDNHNPRDFL ARTQGTAAAR HAAQQNGFEA 60
FAPFAAAVLT AHATGNAGQA TVNTLAGLFI LFRALFIWCY IADKAALRSL MWVGGFVCTV 120
GLFVVAA 127
<212> Type : PRT
35 <211> Length : 127
SequenceName : SEQ ID 234
SequenceDescription :

Sequence

<213> OrganismName : Neisseria meningitidis Z2491
40 <400> PreSequenceString :
MNKIYRIWN SALNAWVAVS ELTRNHTKRA SATVKTAVALA TLLFATVQAN ATDEDEEEEL 60
ESVQRSVVGS IQASMEGSGE LETISLSMTN DSKEFVDPYI VVTLKAGDNL KIKQNTNENT 120
45 NASSFTYSLK KDLTGILNVE TEKLSFGANG KKVNIISDTK GLNFAKETAG TNGDTTVHLN 180
GIGSTLTDL AGSSASHVDA GNQSTHYTRA ASIKDVLNAG WNIKGVKTGS TTGQSENVDF 240
VRTYDTVEFL SADTKTTTVN VESKDNGKRT EVKIGAKTSV IKEKDGLVT GKKGENGSS 300
TDEGEGLVTA KEVIDAVNKA GWRMKTITAN GQTGQADKFE TVTSGTNVTF ASGKGTTATV 360
SKDDQGNITV MYDVNVGDAL NVNQLQNSGW NLDSKAVAGS SGKVISGNVS PSKGKMDTV 420
50 NINAGNNIEI SRNGKNIDIA TSMAPQFSSV SLGAGADAPT LSVDEGALN VGSKDANKPV 480
RITNVAPGVK EGDVTNVAQL KGVAQNLLNR IDNVDGNARA GIAQAIATAG LVQAYLPGKS 540
MMAIGGGTYR GEAGYAIGYS SISDGGNWII KGTASGNSRG HFGASASVGY QW 592
<212> Type : PRT
55 <211> Length : 592
SequenceName : SEQ ID 235
SequenceDescription :

Sequence

<213> OrganismName : Neisseria meningitidis Z2491
60 <400> PreSequenceString :
MLLAEGQKSA VTEYYLNHGT WPSNNSDAGV ASTATDIKKG YVKEVKVEKG VITATMLSSG 60
VNNEIKGKKL SLWAKRQAGS VKWFCGQVE RAANNAANDA VTAATANGNG KIDTKHLPST 120
65 CRDAASAVCI ETPPTAFYKN T 141
<212> Type : PRT
<211> Length : 141

SequenceName : SEQ ID 236
SequenceDescription :

5 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKTDDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN YQYYRDFAEEN 60
KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG VAALVGDQYI VSVAHNGGYN 120
10 NVDFGAEGRN PDQHRFSYQI VKRNNYKPDN SHPYNGDYHM PRLHKFVTD AEPVEMTSDMR 180
GNTYSDKEY PERVRIGSGH HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVVSLSG 240
DVRHANDYGP MPIAGAAGDS GSPMFIYDKT NNNKWLNGVL QTGYPYSGRE NGFQLIRKDW 300
FYDDIYRGDT HTVFFEPSRN GHFSFTSNNN GTGTVTETNE KVSNNPKLVQ TVRLFDESIN 360
ETDKEPVYAA GGVNQYRPRN NNGENLSFID YGNGKLILSN NINQAGAGLY FEGDFTVSPE 420
15 NNETWQAGV HISEDSTVTW KVVGVANDRL SKIGKGTLLV QAKGENQGGI SVGDGTVILD 480
QQADDKGGKQ AFSEIGLVSG RGTVQLNADN QFNPDKLYFG FRGGRLDLNG HSLSFHRIQN 540
TDEGAMIVNH NATTTSTVTI TGNEISITPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR 600
LNLVYQPAAE DRTLLLSGGT NLNGNITQTN GKLFESGRPT PHAYNHLGSG WSKMEGIPQG 660
EIVWDNDWIN RTFKAENFHI QGGQAVISR N VAKVEGDWHL SNHAQAVFGV APHQSHITCT 720
20 RSDWTGLTNC VEKTTTDDKV IASLTKTDIS GNVSLADHAH LNLTLGLATLN GNLSANGDTR 780
YTVSHNATQN GNLSLVGNAQ ATFNQATLNG NTSASGNASF NLSNNAAQNG SLTLSDNAKA 840
NVSHSALNGN VSLADKAVFH FENSRTGQL SGSKDTALHL KDSEWTLPSG TELGNLNLNDN 900
ATITLNSAYR HDAAGAQTGS VSDTPRRRSR RSLLSVTPPT SVESRFNTLT VNGKLNGQGT 960
FRFMSELFY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD QLTVEGKDN KPLSENLNFT 1020
25 LQNEHVDAGA WRYQLIRKDG EFRHLNPNVKE QELSDKLGA EAKKQAEKDN AQSLDALIAA 1080
GRDAAEKTES VAEPARQAGG ENVGIMQAE EKKRVQADKD SALAKQREAE TRPATTAFFR 1140
ARRARRDLPO PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR 1200
NAVWTSGIRD TKHYRSQDFR AYRQQTDLRQ IGMQKNLGGG RVGILFSHNR TENTFDDGIG 1260
NSARLAHGAV FGQYGIGRFD IGISTGAGFS SGSLSDGIGG KIRRRVLHYG IQARYRAGFG 1320
30 GFGIEPYIGA TRYFVQKADY RYENVNIATP GLAFNRYRAG IKADYSFKPA QHISITPYLS 1380
LSYTDAASGK VRTRVNTAVL AQDFGKTRSA EWGVNAEIKG FTLSLHAAAA KGPQLEAQHS 1440
AGIKLGYRW 1449

<212> Type : PRT

<211> Length : 1449

35 SequenceName : SEQ ID 237
SequenceDescription :

Sequence

40 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNTLQKGFIL IELMIVIAIV GILAAVALPA YQDYTARAQV SEAILLAEGQ KSAVTEYYLN 60
HGEWPSNNTS AGVASSTDIK GKYVQSVEVK NGVVTATMAS SNVNNEIKGK KLSLWAKRQD 120
GSVKWFCGQP VKRNDTATTN DDVKADTAAN GKQIDTKHLP STCRDAASAG 170

45 <212> Type : PRT

<211> Length : 170

SequenceName : SEQ ID 238

SequenceDescription :

50 Sequence

<213> OrganismName : Neisseria meningitidis Z2491

<400> PreSequenceString :

55 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAQVK DMDLQALHGR 60
KVALYIATMG DQSGSGLTGG RYSIDALIRG EYINSPAVRT DYTYPYETT AETTSGLTGT 120
LTTSLSTLNA PALSRTQSDG SGSKSSLGLN IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF 180
FLRGIDVVSP ANADTDVFIN IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL 240
IKPKTNAFEA AYKENYALWM GPYKVSIGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN 300
SHEGYGYSDE AVRRHRQGP 320

60 <212> Type : PRT

<211> Length : 320

SequenceName : SEQ ID 239

SequenceDescription :

65 Sequence

<213> OrganismName : Neisseria meningitidis Z2491

<400> PreSequenceString :
MRPIFLSFVL FPILITACST PDKSARWENI GTISNGNIHT YINKDSVRKN GNLMIQDKK 60
VVTNLKQERF ANTPAYKTAI AEWEIHCNNK TYRLSSLQLF DTKNTEISTQ NYTASSLRPM 120
SILSGTLTEK QYETVCGKKL 140

5 <212> Type : PRT
<211> Length : 140
SequenceName : SEQ ID 240
SequenceDescription :

10 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKLFITALS ALALSACAGT WEGAKQDTAR NLDKTQAAAE RAAEQTGNAV EKGWDKTKEA 60
15 VKKGGNAVGR GISHLGKIE NATE 84
<212> Type : PRT
<211> Length : 84
SequenceName : SEQ ID 241
SequenceDescription :

20 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
25 MKLLFIPLVL FVAVEHFYIA WLEMTQIPSE KAAETFKLPY EFMEQNRVQT LFGNQGLYNG 60
FLGIGLVWSR FAAPDNAVYG ATVLFLGFVL IAAAWGAFSS GNKGILVKQG LPAFLAAAV 120
LAV 123
<212> Type : PRT
<211> Length : 123
30 SequenceName : SEQ ID 242
SequenceDescription :

Sequence

35 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MASSNVNNEI KDKKLSLWAK RQDGSVKWFC GQPVKRDAAT DADVTADSGN EIDTKHLPST 60
CRDAASAVCT KTPEYYPNHG EWPKNFVIPA QAGIQVCRHG NLSGKKVSPV LSSRFPLSWE 120

40 <212> Type : PRT
<211> Length : 120
SequenceName : SEQ ID 243
SequenceDescription :

45 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MLLAEGQKSA VTEYYLNHGE WPSNNTSAGV ATSTDIKGY VQSVEVKNGV VTATMASSNV 60
50 NNEIKGKKLS LWAKRQDGSV KWFCGQPVKR NDTATTNDDV KADTAANGKQ IDTKHLPSTA 120
STRKSTPN 128
<212> Type : PRT
<211> Length : 128
SequenceName : SEQ ID 244
55 SequenceDescription :

Sequence

<213> OrganismName : Neisseria meningitidis Z2491
60 <400> PreSequenceString :
MPIPFPVLA AAAIAQAFPA FAADPAPQSA QTLNEITVTG THKTQKLGE KIRRKTLDKL 60
LVNDEHDLVR YDPGISVVEG GRAGSNGFTI RGVDKDRVAI NVDGLAQAES RSSEAFQELF 120
GAYGNFNANR NTSEPFNFSE VTITKGADSL KSGSGALGGA VNYQTKSASD YVSEDKPYHL 180
GIKGGSVGKN SQKFSSITAA GRLFGLDALL VYTRRFKGK KNRSTEGDIE IKNDGYVYNP 240
65 TDTGGPSKYL TYVATGVARS QPDPQEWVKN STLFLKGYNF NDQNRIGWIF EDSRTDRFTN 300
ELSNLWTGTT TSAATGDYRH RQDVSYYRRS GVEYKNELEH GPWDSLKLRY DKQRIDMNTW 360
TWDIPKNYDK RGINGEVYHS FRHIRQNTAQ WTADFEKQLD FSKAVWAAQY GLGGGKGDNA 420

NSDYSYFAKL YDPKILASNQ AKITMLIENR SKYKFAYWNN AFHLGGNDRF RLNAGIRYDK 480
NSSSAKDDPK YTTAIRGQIP HLGSERAHAG FSYGTGFDWR FTKHLHLLAK YSTGFRAPTS 540
DETWLLFPHF DFYLVKANPNL KAELAKNWEL GLAGSGKAGN FKLSGFKTKY RDFIELTYMG 600
VSSDDKNNPR YAPLSDGTAL VSSPVWQONQ RSAAWVKGIE FNGTWNLDSI GLPKGLHTGL 660
5 NVSYIKGKAT QNNGKETPIN ALSPWTAVYS LGYDAPSKRW GINAYATRТА AKKPSDTVHS 720
NDDLNNPWPY AKHSKAYTLF DLSAYLNIGK QVTLRAAAYN ITNKQYYTWE SLRSIREFGT 780
VNRVDNKTHA GIORFTSPGR SYNFTIEAKF 810
<212> Type : PRT
<211> Length : 810
10 SequenceName : SEQ ID 245
SequenceDescription :

Sequence

15 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKKSLIALTL AALPVAAMAD VTLYGTIKTG VETSRSEVHN GGQVSVETG TGIVDLGSKI 60
GFKGQEDLGN GLKAIWQVEQ KASIAGTDSG WGNRQSFGL KGGFGKLRVG RLNSVLKDTG 120
DINPWSKSD YLGVNKIAEP EARLISVRYD SPEFAGLSGS VQYALNDNVG RHNSSEYHAG 180
20 FNYKNGGFFV QYGGAYKRHO DVDDVKIEKY QIHRLVSGYD NDALYASVAV QQQDAKLVED 240
NSHNSQTEVA ATLAYRFGNV TPRVSYAHGF KGSVDDAKRD NTYDQVVVGA EYDFSKRTSA 300
LVSAGWLQEG KGENKFVATA GGVGLRHKF 329
<212> Type : PRT
<211> Length : 329
25 SequenceName : SEQ ID 246
SequenceDescription :

Sequence

30 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKTLILLIPL VLTACGTLTG IPAHHGGKRF AVEQELVAAS SRAAVKEMDL SALKGRKAAL 60
YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY PAYDTTATTK SDALSSVTTS 120
TSLLNAPAAA LTKNSGRKGE RSAGLSVNGT GDYRNETLLA NPRDVSFLTIN LIQTVFYLRG 180
35 IEVVPPEYAD TDVFVTVDVF GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK 240
TAAYESQYQE QYALWMGPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNGKKP 300
DVGNEVIRRR KGG 313
<212> Type : PRT
<211> Length : 313
40 SequenceName : SEQ ID 247
SequenceDescription :

Sequence

45 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKTLISILPV AILLGGCAAG GGNTFGSLDG GTGMGGSIVK MAVESQCRAE LNKRESEWRLT 60
ALAMSAEKQA EWENKICACV AQEAPNQLTG NDVMQMLDPS TRNQALAAIT AKTVSACFKH 120
LYR 123
50 <212> Type : PRT
<211> Length : 123
SequenceName : SEQ ID 248
SequenceDescription :

55 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNPLIHQAKE SSMQTRILSA VLLAFSTAAF AGGAFTLQFD NPSEDGGFTQ NQILSAPYGF 60
60 GCSGGNASPA LSWKNPPAGT KSFVLTVYDK DAPTGLGWMH WVADIPADV RRRNATSLQL 120
SRCASIADDQ SAAISAVISL QICRIRLTPS YTAKPMPSCC NHANTPQSAA SAALCGTSSS 180
VSTAAA 186
<212> Type : PRT
<211> Length : 186
65 SequenceName : SEQ ID 249
SequenceDescription :

Sequence

<213> OrganismName : Neisseria meningitidis Z2491

<400> PreSequenceString :

5 MNKTLKRRVF RHTALYAAIL MFSHTGGGGG AMAQTRQYAI IMNERNQPEV QWNGSYSIKD 60
KDRKREYTHH NHQQGGSSVS FNNSEDELVS QSGTAVFGTA TYLPPYGKVS GFDAAALKER 120
NNAVDWIHTT HPGLIGYSYD GVVCRSATDC PKLVYKTRFS FDNPD LAKTG GGLDKHTEPS 180
RDNSPIYK LK DHPWLGVSFN LGAEGIAKNG KTINKLVSSF NEKNSNNNLV YTTEGRDISL 240
GNWQRETTAM AYYLNAKLHL LDKKQIQNIT DKT VQLGV LK PSIDVTRNT GTAGILSYWA 300
10 KWDIKDTGQI PVKLSLTQVK AGRCVNKDNP NKNTKTSSPA LTAPALWFGA QDGKAEMYS 360
ASVSTYPDSS SSRIFLQNLK RKTDTSRPGR YSLATLNKSD IESREPSFTS RQTVIRLDGG 420
VQQIKLDRNN TEVTGFNGND GKNDTFGIVS EGSFMPDASE WKKVLLPWT V RAFNYDGRFN 480
TVNKEENNGK PKYSQKYRSR NNGKHERNLG DIVNSPIVAV GEYLATSAND GMVHIFKQSG 540
GDKRSYNLKL SYIPGTMPRK DIESKDSTLA KELRAFAEKG YVCDRYGVDG GFVLRRI1DD 600
15 QDKQKHFFMF GAMGLGGRGA YALDLTKADD NDPTKASLFD VKDNGNNGNN GNNRVELGYT 660
VGTPQIGKTH NGKYAAFLAS GYATKQIDSG ENKTALYVYD LESNNGTLIR KIEVTDGKGG 720
LSSPTLV D K LDGTVDIAYA GDRGGKMYRF DLSGNNPNSW TVRTIFQGTK PITSAPAI SQ 780
LKDKRVVIFG TGS D LSEDDV LSTDEQHIY G IFDNDTNTGT AQEGLGKGLL EQKLSEENKT 840
LFLTDYKRSD GSGDKGWVVK LKDGQRVTVK PTVVLRTAFV TIHKYTGNDK CGAETAILGI 900
20 NTADGGK LTK KSARPIVPAA NSKVAQYSGD KKTSSGKSIP IGCMEKDGGT VCPNGYVYDK 960
PVNVRYLDEK KTDGFSTTAD GDAGSGSTFK EGKKPARNNR CFSGKGVRTL LMNDLDSL DI 1020
TGPMCGMKRI SWREVFY 1037

<212> Type : PRT

<211> Length : 1037

25 SequenceName : SEQ ID 250

SequenceDescription :

Sequence

30 <213> OrganismName : Neisseria meningitidis Z2491

<400> PreSequenceString :

MKHPKLTLIA ALLTTAATAA PLPVVTSFSI LGDVAQIGG ERVSIQSLVG ANQDTHAYHM 60
TSGDIKKIRS AKLV LINGLG LEAADIQRAV KQSKVSYAEA TKGIQPLKAE EEGGHHHDHD 120
HHDHHDHEGH HHDHGEYDPH VWN D PVLMSA YAQNVAEALI KADPEGKVYY QORLGNYQMQ 180
35 LKKLHSDAQA AFNAVPAKR KVL TGHDAFS YMGKRYHIEF IAPQGV SSEA EPSAKQVA AI 240
IRQIKREGIK AVFTENIKDT RMVDRIAKET GVNVS GKLYS DALGNAPADT YIGMYRHN IK 300
ALTNAMKQ 308

<212> Type : PRT

<211> Length : 308

40 SequenceName : SEQ ID 251

SequenceDescription :

Sequence

45 <213> OrganismName : Streptococcus pyogenes MGAS8232

<400> PreSequenceString :

MKKRILSAVL VSGVTLGAAT TVGAEDLSTK IAKQDSIISN LTTEQKAAQN QVSALQAQVS 60
SLQSEQDKLT ARNTELEALS KRFEQEIKAL TSQIVARNEK LKNQARSAYK NNETSGYINA 120
LLNSKSISDV VNRLVAINRA VSANAKLLEQ QKADKVSLEE KQANQTAIN TIAANMAMAE 180
50 ENQNTLRTQQ ANLEAATANL ALQLASATED KANLVAQKEA AEKAAAEALA QEQA AKVKAQ 240
EQAAQQAASV EAKSAITPA PQATPAAQSS NAIEPAALTA PAAPSARPQT SYDSSNTYPV 300
GQCTWGA KSL APWAGNNWGN GGQWAYSQA AGYRTGSTPM VGAI AVWNDG GYGHVAVVVE 360
VQSASSIRVM ESNYSGRQYI ADHRGWFNPT GVTFIYPH 398

<212> Type : PRT

55 <211> Length : 398

SequenceName : SEQ ID 252

SequenceDescription :

Sequence

60 <213> OrganismName : Streptococcus pyogenes MGAS8232

<400> PreSequenceString :

MITIKNP KIL KWLKYVLSAI LSLIILVIII GGLLFTFYIS SAPKLSEAQL KSTNSSLVYD 60
GNNNLIADLG SEKRENTAD SIPINLVNAI TSIEDKRFFN HRGVDLYRIF GAAFHNLT SQ 120
65 TTQGGSTLDQ QLIKLAYFST NESDQTLKRK AQEVWLALQM ERKYTKQEIL TFYINKVYMG 180
NGNYGMLTAA KSYYGKDLKD LSYAQLALLA GIPQAPSQYD PYLHPEAAQN RRNVVLQ QMY 240
MEKHLTKAEY ETAIATPVAE GLQSLQQRST YPKYMDNYLK QVIEEVKKET NKDIFTAGLK 300

VYTNIIIPDAQ QTLYNIYHSG DYVYYPDQDF QVASTIVDVT NGHVIQQLGG RNQDENVSFG 360
TNQAVLTDRD WGSTMKPITA YAPAIESGVY TSTAQSTNDS VYYWPGTTTQ LFNWDLRYNG 420
WMTIQAAIML SRNVPVAVRAL EAAGLDYARS FLSSLGINYP EMHYSNAISS NNSSSDKKYG 480
ASSEKMAAAY AAFANGGIYH KPRYVNVKVEF SDGTSKTFDE KGKRAMKETT AYMMTDMLKT 540
5 VLTGTYGTAA AIPGVAQAGK TGTSNYTDEE LAKIGEKYGL YPDYVGT LAP DENFVGFTKR 600
YAMAVWTGYK NRLTPVYGSS LEIASDVYRS MMTYLTNGYS EDWTMPNGLY RSGGFLYLSG 660
TYASNTDYN SVYNNLYSMN TTTASSQTTS DDTSSSNDTS NSTNTDNNGS HPSTDDKKT 720
H 721
<212> Type : PRT
10 <211> Length : 721
SequenceName : SEQ ID 253
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MIITKKSFLV TSVALSLAPL VTAQAQEWTP RSVTEIKSEL VLVDNVFTYT VKYGD TLSTI 60
AEAMGIDVHV LGDINHIANI DLIFPDILT ANYNQHGQAT TLT VQAPASS PASVSHVPSS 120
20 EPLPQASATS QSTVPMAPSA TPSDVPTTPL ASAKPDSFVT ASSELTSSSTN DVSTELSSSES 180
QKQPEVSQEA VPTPKAAETT EVEPKTDISE DPTSANRPVP NESASEEASS AAPAQAPAEK 240
EETSQMLTAP AAQKAVADTT SVATSNGLSY APNHAYNPMN AGLQPQTAAF KEEVASAFGI 300
TSFSGYRPGD PGDHGKGLAI DFMVPVSSTL GDQVAQY AID HMAERGISYV IWKQRFYAPF 360
ASIYGPAYTW NPMPDRGSIT ENHYDHVHVS FNA 393
25 <212> Type : PRT
<211> Length : 393
SequenceName : SEQ ID 254
SequenceDescription :

Sequence
30 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKKILLMMS LISVFFAWQL TQAKQVLAEG KVKVVTTFYP VYEFTKGVIG NDGDVSM LMK 60
35 AGTEPHDFEP STKDIKKIQD ADAFVYMDN METWVSDVKK SLTSKKVTIV KGTGNM L LVA 120
GAGHDHHHED ADKKHEHNKH SEEGHNHAFD PHVWLSPYRS ITVENIRDS LSKAYPEKAE 180
NFKANAATYI EKLKELDKDY TAALSDAKQK SFVTQHAAFG YMALDYGLNQ ISINGVTPDA 240
EPSAKRIATL SKYVKKYGIK YIYFEENASS KVAKTLAKEA GVKAAVLSPL EGLTKKEMKA 300
GQDYFTVMRK NLET LRLTTD VAGKEILPEK DTTKTVYNGY FKDKVKDRQ LSDWSGSWQS 360
40 VYPYLQDGT L DQVWDYKAKK SKGKMTAAEY KDYYTTGYKT DVEQIKINGK KKTMTFVRNG 420
EKKTFTYTYA GKEILTYPKG NRGVRFMFEA KEPNAGEFKY VQFSDHAIAP EKAEHFHL YW 480
GGDSQEK LHK ELEHWPTYYG SDLSGREIAQ EINAH 515
<212> Type : PRT
<211> Length : 515
45 SequenceName : SEQ ID 255
SequenceDescription :

Sequence
50 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKFHRFLVS GVILLGFNGL VPTMPSTLIS QQENLVHAAV LGDNYP SKWK KNGIDSWNM 60
YIRQCTSF AA FRLSSANGFQ LPKGYGNACT WGHIKNQGY PVNKTPSIGA IAWFDKNAYQ 120
SNAAYDHVAV VADIRGDTV T IEEYNYNAGQ GPERYHKRQI PKSQVSGYIH FKDLSSQTS H 180
55 SYPRQLKHIS QASFDPSGT Y HFTTRLPVKG QTSIDSPDLA YYEAGQSVYY DKVV TAGGYT 240
WLSYLSFSGN RRYIPIKEPA QSVVQNDNTK PSIKVGDTV T FPGVFRVDQL VNNLI VNKEL 300
AGGDPTPLNW IDPTPLDET D NQGKVLGNQI LRVGEYFTVT GSYKVLKIDQ PSNGIYVQIG 360
SRGTWVNADK ANKL 374
<212> Type : PRT
60 <211> Length : 374
SequenceName : SEQ ID 256
SequenceDescription :

Sequence
65 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :

MLKFTSNILA TSVAETTQVA PGGCCCCCTT CCFSIATGSG NSQGGSGSYT PGK

<212> Type : PRT
<211> Length : 53
5 SequenceName : SEQ ID 257
 SequenceDescription :

Sequence

10 <213> OrganismName : Streptococcus pyogenes MGAS8232
 <400> PreSequenceString :
MGESYSVEAV LTAVDKTFGK TLQSAIRSIE GLEKRSTGFS SVSQKASSMF KSMLGANLAG 60
QAISAMTRTV SSGLGSMLE MNSSAKAWKT FDANLADIGF GKKQILAVKT AMQDYATKTI 120
YSASDMASTY AQLAAVGVKD TGKLVKAFGG LAASAENPKQ AMKSISQQMT QAVGRPTVAW 180
15 QDFRIMLEQT PAGMAKVAKS MGKNLDELVA DIQAGRVKTS DFLEAVKKAG NDKSFQKMAT 240
EFKTVDAQID GMREGLSNKL QPAFEKVNQF GIRAIEAIGK QLDKVDVFSKF ASNLGKFLEG 300
INIDKIVSNI SSAVSSVTSK VKEFWDGFKQ TGAISAFSGA LQSVWGALKN VASAMSGGNW 360
KTFGATVGGI VKHVSNFAKA VSDVLGKMDP GRLRSWIATF AAVAGGFKLF EKLTGQSVIG 420
SFLDKIGSKF GLFGNKAKEG TDKASNGARR SGGIISQIFS GLGNIVKSAG TAISTAAGKI 480
20 GVGIKTALSG IPPYH 495
 <212> Type : PRT
 <211> Length : 495
 SequenceName : SEQ ID 258
 SequenceDescription :

25 Sequence

30 <213> OrganismName : Streptococcus pyogenes MGAS8232
 <400> PreSequenceString :
MKKGFFLMVM VVSLVMIAGC DKSANPKQPT QGMSVVTSFY PMYAMTKEVS GDLNDVRMIQ 60
SGAGIHSFEP SVNDVAAIYD ADLFVYHSHT LEAWARLDLP NLKKSVDVDF EASKPLTLDR 120
VKGLEDM EVT QGIDPATLYD PHTWTDVLA GEEAVNIAKE LGRLDPKHKD SYTKNAKAFK 180
KEAEQLTEFY TQKFKKVRSK TQVHTAFS YLAKRFGKQ LGISGISPEQ EPSRQLKEI 240
QDFVKEYNVK TIFAEDNVNP KIAHAIAKST GAKVKTLSPLEAAPSGNKTY LENLRANLEV 300
35 LYQQLK 306
 <212> Type : PRT
 <211> Length : 306
 SequenceName : SEQ ID 259
 SequenceDescription :

40 Sequence

45 <213> OrganismName : Streptococcus pyogenes MGAS8232
 <400> PreSequenceString :
MEKKQRFSLR KYKSGTFSVL IGSVFLMMTT TVAADELSTM SEPTITNHTQ QQAQHLTNTE 60
LSSAESKSQD TSQITPKTNR EKEQPQGLVS EPTTTELADT DAAPMANTGP DATQKSASLP 120
PVNTDVHDVW KTKGAWDKGY KGQGVAVI DTGIDPAHQS MRISDVSTAK VKSKEDMLAR 180
QKAAGINYGS WINDKVFAH NYVENSNDIK ENQFEDFDED WENFEFDAEA EPKAIKKHKI 240
YRPQSTQAPK ETVIKTEETD GSHDIDWTQT DDDTKYESHG MHVTGIVAGN SKEAAATGER 300
50 FLGIAPEAQV MFMRVFANDV MGSAESLFK AIEDAVALGA DVINLSLGTA NGAQLSGSKP 360
LMEAIEKAKK AGVSVVVAAG NERVYGSDDH DPLAINPDYG LVGSPSTGRT PTSVAAINSK 420
WVIQRLMTVK ELENRADLNH GKAIYSESVD FKNIKDSLGY DKSHQFAYVK ESTDAGYKAQ 480
DVKDKIALIE RDPNKTYDEM IALAKKHGAL GVLIFFNNKPG QSNRSMRLTA NGMGIPSAFI 540
SHEFGKAMSQ LNGNGTGSLE FDSVVS KAPS QKGNEMNHFS NWGLTSDGYL KPDITAPGGD 600
55 IYSTYNDNHY GSQTGTSMAS PQIAGASLLV KQYLEKTQPN LPKEKIADIV KNLLMSNAQI 660
HVNPEKTTT SPRQAGALL NIDGAVTSGL YVTGKDNYGS ISLGNITDTM TFDVTVHNLIS 720
NKDKTLRYDT ELLTDHVDPO KGRFTLTSRS LKTYQGGEVT VPANGKVTVR VTMDVSQFTK 780
ELTKQMSNGY YLEGFVRFRD SQDDQLNRVN IPFVGFKGQF ENLVAEESI YRLKSQKGTG 840
FYFDESGPKD DIYVGKHFTG LVTLGSETNV STKTISDNGL HTLGTFTKNAD GKFILEKNAQ 900
60 GNPVLAISPN GDNNQDFAAF KGVFLRKYQG LKASVYHASD KEHKNPLWVS PESFKGDKNF 960
NSDIRFAKST TLLGTAFSGK SLTGAEPLDG YYHYVVSYP DVVGAKRQEM TFDMLDRQK 1020
PVLSQLATFD ETRNFKEPEL KDRGLAGVRK DSVFYLERK NKPYTVTIND SYKYVSVEDN 1080
KTFVERQADG SFILPLDKAK LGDFYMYVED FAGNVAIAKL GDHLPQTLGK TPIKLLKLTG 1140
NYQTKETLKD NLEMTQSDTG LVTNQAQLAV VHRNQPSQL TKMNQDFFIS PNEDGNKDFV 1200
65 AFKGLKNNVY NDLTVNVYAK DDHQKQTPIW SSQAGASASA IESTAWYGIT ARGSKVMPGD 1260
YQYVVYTYRDE HGKEHQKQYT ISVNDKKPMI TQGRFDITING VDHFTPDKTK ALGSSGIVRE 1320
EVFYLAKKNG RKFDVTEGKD GITVSDNKMY IPKNPDGSYT ISKRDPVTLIS DYYYLVEDRA 1380

GNVSFATLRD LKAVGKDKAV VNFGLDLPVP EDKQIVNFTY LVRDADGKPI ENLEYNNNSG 1440
NSLILPYGKY TVELLTYDTN AAKLESDKIV SFTLSADNNF QQVTFKMTML ATSQITAHFD 1500
HLLPEGSRVS LKTAQGQLIP LEQSLYVPA YGKTVQEGTY EVVVS LPKG Y RIEGNTKVNT 1560
LPNEVHELSE RLVKVG DAST STGDHKVMSK NNSQALTAFA TPTKTTTSAT AKALPSAGEK 1620
5 MGLKLRIVGL VLLGLTCVFS RKKSTKD 1647
<212> Type : PRT
<211> Length : 1647
SequenceName : SEQ ID 260
SequenceDescription :

10 Sequence

<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :

15 MMRS LFSGVS GMQNHQTRMD VIGNNVANVN TTGFKRGRVN FQDLISQQLS AAARPNEEVG 60
GVNPKEVGLG VLIASIDTVH TQALQTTGI NTDVSIQSGG FFVLKSGEKT FFTRAGAFGV 120
DNAGTLVNPA NGMRVQGWMA QDVAGERLIN SSAQTQDLVI PIGQKIDAQQ TSTVHYACNL 180
DKRLPELAAD ANEADVRSKST WTTDFQVYDS FGQOHTLQIN FSRVPGTNNQ WQATVAVDPG 240
TEVDITQTRVG VGTSDGAANT FIVNFDNFGH LASVTDTAGN VTGPTGQVLL EASYDVVGAN 300
20 PDDAGQVTRH AFTLNLGEIG TARNTITQFA ERSTTKAYRQ DGYAMGYLEN FKIDQSGVIT 360
GVYSNGVSQD IGQLALAGFA NQGGLEKAGE NTYVQSNNSG IANISTSGVM GKGLIAGTL 420
EMSNVDLTDQ FTDMIITQKG FQAGAKTIQT SDTMLDTVLS LKR 463
<212> Type : PRT
<211> Length : 463
SequenceName : SEQ ID 261
SequenceDescription :

25 Sequence

<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :

30 MGC MRWGSVL CVVVG VGASG GVLGQEFSPK LTGSATLEWG ISYKG VGSHE QAPGAVMGT 60
GPYNLKHGFR TTNTVGVSFP LVMRTTHTRR GQHPALYABL KVADLQADLS QGKAGFAVKR 120
KGKVEATLHC YGAYLTIGKN PTFLT NFARL WKPWVTAQYQ EDAVQYAPGE GGLGGKVG YR 180
35 AQDIGGSGVS LDVGFLSFAS NGAWDSTDPT HSKYGF GADL KLMYARAGHP LCTVELASNV 240
TLEDGYLIGA QKDANNQNKD KLLWNVGGRL TLEPGAGFRF SFALDAGNQH QSAQDFQNR T 300
QRAQSELTAL SNNLFQGESQ KQEA WVTQVV QQATQTVTAG VRSALSRGT TYINALEAVQ 360
PNPAKPTGKV VQNLHTPQGS PPNLPPLPAL PAFSLMGQVL LQYDAEQVVK GFEQVQTQIV 420
TEINQKVQAA VAKNNANMQA VGGSLGDTAR MVGEALIKQQ LSRKQNSILT MVSQDEVKQ 480
40 DLADLVPMMR TEITAFFASV QQHIT EEVKK KTDALNAGQQ IRQAIQNLRA SAWRAFLMGV 540
SAVCLYLD TY NVAFDALFTA QWKWLSSGIY FATAPANVFG TRVLDNTIAS CGDFAGFLKL 600
ETKSGDPYTH LLTGLDAGVE TRVYIPLTHD LYKNNNGNPL PSGGSSGHIG LPVVGKAWCS 660
YRIPVQDYGW VKPSVTVHAS TNRAHLNAPA AGGAVGATYL TKEYCAQLRA GISASLIEKT 720
VFSLDWEQGM LSDVPYLLVS ECLTQGIGRI VCGVTL SW 758
45 <212> Type : PRT
<211> Length : 758
SequenceName : SEQ ID 262
SequenceDescription :

50 Sequence

<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :

55 MGRQVMQAGV LAGMVCAASG YAGVLT PQVS GTAQLQWGIA FQKNPRTGPG KHTHGFRTTN 60
SLTISLPLVS KHTHTRGEA RSGVWAQLQL KD LAVE LASS KSSTALSFTK PTASFQATLH 120
CYGAYLTVGT SPSCVNFQA LWKPFVTRAY SEK DTRYAPG FSGSGAKLGY QAHNVGNSGV 180
DVDIGFLSFL SNGAWDSTD T HSKYGF GAD ATLSYGVDRQ RLLTLELAGN ATLDQNYVKG 240
TEDSKNENKT ALLWGVGGRL TLEPGAGFRF SFALDAGNQH QSNHAQTQE RAILKAREVF 300
RRVEGKLVQN LPNIMMPPI TEQTTLIEMV GLAALIAEGT LGS AIQT VLA AGALAALVSQ 360
60 LVPNIEQGV R DVFRSSDPRV VTAKLLAFLE RAPMNALNID ALLRMQWKWL SSGIYFATAG 420
TNIFGKRVFA TTRAHYFDFA GFLKLETKSG DPYTHLLTGL NAGVEARVYI PLTYIRYRNN 480
GGYELNGAVP PGTINMPILG KAWCSYRIPL GSHAWLAPHT SVLGTTNRFN IINPAGNLLN 540
ERALQYQVGL TFSPFEKVEL SAQWEQGVLA DAPYMGIAES IWSERHFGTL VCGMKVTW 598
65 <212> Type : PRT
<211> Length : 598
SequenceName : SEQ ID 263

SequenceDescription :

Sequence

5 <213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :
MGRQVMQAGV LAGMVCAASG YAGVLTQVVS GTAQLQWGIA FQKNPRTGPG KHTHGFRTTN 60
SLTISLPLVS KHTHTRRGEA RSGVWAQLQL KDLAVELASS KSSTALSFTK PTASFQATLH 120
CYGAYLTVGT SPSCVVNFAQ LWKPFVTRAY SEKDRYAPG FSGSGAKLGY QAHNVGNSGV 180
10 DVDIGFLSFL SNGAWDSTD TTHSKYGFAD ATLSYGVDRQ RLLTLELAGN ATLDQNYVKG 240
TEDSKNENKT ALLWGVGGRL TLEPGAGFRF SFALDAGNQH QSNHAQTQE RAILKAREVF 300
RRVEGKLVQN LPNIMPPGI TEQTTLIEMV GLAALIAEGT LGSATQTVLA AGALAALVSQ 360
LVPNIEQGVV DFRSSDPRV VTAKLLAFLE RAPMNALNID ALLRMQWKWL SSGIYFATAG 420
TNIFGKRVFA TTRAHYFDFA GFLKLETSG DPYTHLLTGL NAGVEARVYI PLTYIRYRNN 480
15 GGYELNGAVP PGTINMPILG KAWCSYRIPL GSHAWLAPHT SVLGTNTRFN IINPAGNLLN 540
ERALQYQVGL TFSPEKVEL SAQWEQGVLA DAPYMGIAES IWSERHFGTL VCGMKVTW 598

<212> Type : PRT

<211> Length : 598

20 SequenceName : SEQ ID 264
SequenceDescription :

Sequence

25 <213> OrganismName : SARS coronavirus Frankfurt 1
<400> PreSequenceString :
MFIFLLFLTL TSGSDDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
30 HLREFVFNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSLVYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNYKY RYL RHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
35 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
40 GFNFSQLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
TVLPPLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
45 GTSWFITQRN FFSPQIITD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVDFGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT

<211> Length : 1255

50 SequenceName : SEQ ID 265
SequenceDescription :

Sequence

55 <213> OrganismName : SARS coronavirus HSR 1
<400> PreSequenceString :
MFIFLLFLTL TSGSDDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
60 TNVVIRACNF ELCDNPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSLVYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
65 LAWNTRNIDA TSTGNYNYKY RYL RHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600

5 VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NOKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
10 HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT

<211> Length : 1255

15 SequenceName : SEQ ID 266
SequenceDescription :

Sequence

20 <213> OrganismName : SARS coronavirus ZJ01
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFSGTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQTH T MIFDNFNCT FEYISDAFSL DVSEKSGNFK 180
25 HLREFVFNK DGFLYVYKGY QPIDVVRDL SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
30 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
35 GFNFSQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NOKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
40 GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT

45 <211> Length : 1255

SequenceName : SEQ ID 267
SequenceDescription :

Sequence

50 <213> OrganismName : SARS coronavirus TW1
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFSGTMNNK SQSVIIINNS 120
55 TNVVIRACNF ELCDNPFFAV SKPMGTQTH T MIFDNFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFNK DGFLYVYKGY QPIDVVRDL SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
60 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
65 NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900

5 NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT
<211> Length : 1255
10 SequenceName : SEQ ID 268
SequenceDescription :

Sequence

15 <213> OrganismName : SARS coronavirus CUHK-Su10
<400> PreSequenceString :

20 MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
25 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
30 GFNFSQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
35 GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT
40 <211> Length : 1255
SequenceName : SEQ ID 269
SequenceDescription :

Sequence

45 <213> OrganismName : SARS coronavirus Urbani
<400> PreSequenceString :

50 MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
55 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
60 NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
65 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200

GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT
<211> Length : 1255
5 SequenceName : SEQ ID 270
 SequenceDescription :

Sequence

10 <213> OrganismName : SARS coronavirus
 <400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
15 HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
 AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
 QTSNFRVVPs GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
 FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
20 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
 SSKRFQPFQ Q FGRDVSDFTD SVRDPKTSEI LDISPCAFGG VSVITPGTNA SSEVAVLYQD 600
 VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
 HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
 NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
25 GFNFSQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
 TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
 NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
 DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
30 GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
 HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
 GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT
35 <211> Length : 1255
 SequenceName : SEQ ID 271
 SequenceDescription :

Sequence

40 <213> OrganismName : SARS coronavirus Tor2
 <400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
45 TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
 HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
 AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
 QTSNFRVVPs GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
 FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
50 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
 SSKRFQPFQ Q FGRDVSDFTD SVRDPKTSEI LDISPCAFGG VSVITPGTNA SSEVAVLYQD 600
 VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
 HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
55 NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
 GFNFSQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
 TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
 NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
 DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
60 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
 GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
 HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
 GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

65 <212> Type : PRT
 <211> Length : 1255
 SequenceName : SEQ ID 272

SequenceDescription :

Sequence

5 <213> OrganismName : SARS coronavirus GD01
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFSGTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQTH T MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
10 HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFLP 240
AQDTWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPs RDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKRISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNKY RYLRHGKLRP FERDISNVFP SPDGKPCTPP ALNCYWPLND 480
15 YGFYTTTGIG YQPYRVVLS YELNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQO FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYEC DI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKDFG 780
20 GFNFSQILPD PLKSTKRSFI EDLLFNKVT L ADAGFMKQYG ECLGDINARD LICAQKFNG L 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
25 GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD IGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYE Q YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT

30 <211> Length : 1255

SequenceName : SEQ ID 273

SequenceDescription :

Sequence

35 <213> OrganismName : SARS coronavirus CUHK-W1
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFSGTMNNK SQSVIIINNS 120
40 TNVVIRACNF ELCDNPFFAV SKPMGTQTH T MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDTWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPs GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
45 LAWNTRNIDA TSTGNYNKY RYLRHGKLRP FERDISNVFP SPDGKPCTPP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQO FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYEC DI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
50 NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
GFNFSQILPD PLKPTKRSFI EDLLFNKVT L ADAGFMKQYG ECLGDINARD LICAQKFNG L 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
55 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD IGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYE Q YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT

60 <211> Length : 1255

SequenceName : SEQ ID 274

SequenceDescription :

Sequence

<213> OrganismName : SARS coronavirus BJ01

<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQTH T MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
5 H LREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDTWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVP S GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVADIYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
10 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSPGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
15 GFNFSQILPD PLKPTKRSFI EDLLFNKVT L ADAGFMKQYG ECLGDINARD LICAQKFNG L 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
20 GTSWFITQRN FFSPQIITD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT
25 <211> Length : 1255
SequenceName : SEQ ID 275
SequenceDescription :

Sequence
30 -----
<213> OrganismName : SARS coronavirus
<400> PreSequenceString :
SGFRKMAFPS GKVEGCMVQV TCGTTTLNGL WLDDTVYCPR HVICTAEDML NPNYEDLLIR 60
KSNHSFLVQA GNVQLRVIGH SMQNCLRLRK VDTSNPKTPK YKFVRIQPGQ TFSVLACYNG 120
35 SPSGVYQCAM RPNHTIKGSF LNGSCGSVGF NIDYDCVSFC YMHMELPTG VHAGTDLEGK 180
FYGPFVDRQT AQAAGTDTTI TLNVLAWLYA AVINGDRWFL NRFTTTLNDF NLVAMKYNYE 240
PLTQDHVDIL GPLSAQTGIA VLDMCAALKE LLQNGMNGRT ILGSTILEDE FTFPDVVRQC 300
SGVTFQ 306
<212> Type : PRT
40 <211> Length : 306
SequenceName : SEQ ID 276
SequenceDescription :

Sequence
45 -----
<213> OrganismName : SARS coronavirus
<400> PreSequenceString :
AIASEFSSLP SYAAYATAQE AYEQAVANGD SEVVLKLLKK SLNVAKSEFD RDAAMQRKLE 60
KMADQAMTQM YKQARSEDKR AKVTSAMQTM LFTMLRKLND DALNNIINNA RDGCVPLNII 120
50 PLTTAAKLMV VVPDYGTYN TCDGNTFTYA SALWEIQQVV DADSKIVQLS EINMDNSPNL 180
AWPLIVTALR ANSAVKLQ 198
<212> Type : PRT
<211> Length : 198
SequenceName : SEQ ID 277
55 SequenceDescription :

Sequence

<213> OrganismName : SARS coronavirus
60 <400> PreSequenceString :
AGNATEVPAN STVLSFCAFA VDPKAYKDY LASGGQPITN CVKMLCTHTG TGQAITVTPE 60
ANMDQESFGG ASCCLYCRCH IDHPNPKGFC DLKGKYVQIP TTCANDPVGF TLRNTVCTVC 120
GMWKGYGCSC DQLREPLMQ 139
<212> Type : PRT
65 <211> Length : 139
SequenceName : SEQ ID 278
SequenceDescription :

Sequence

<213> OrganismName : SARS coronavirus
5 <400> PreSequenceString :
NNELSPVALR QMSCAAGTTQ TACTDDNALA YYNSKGGRF VLALLSDHQD LKWARFPKSD 60
GTGTIYTELE PPCRFVTDTP KGPVKYLYF IKGLNNLNRG MVLGSLAATV RLQ 113

<212> Type : PRT
10 <211> Length : 113
SequenceName : SEQ ID 279
SequenceDescription :

15 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
20 MNKIFKVIWN PATGNYTVTS ETAKSRGKKS GRSKLLISAL VAGGMLSSFG ALANAGNDNG 60
QGVDYGSOSA GDGWVAIGKG AKANTFMNTS GSSTAVGYDA IAEGQYSSAI GSKTHAIGGA 120
SMAFGVSAIS EGDRSIALGA SSYSLGQYSM ALGRYSKALG KLSIAMGDSS KAEGANAIAL 180
GNATKATEIM SIALGDTANA SKAYSMALGA SSVASEENAI AIGAETEAEE NATAIGNNAK 240
AKGTNSMAMG FGSLADKVNT IALNGSQAL ADNAIAIGQG NKADGVDAIA LGNGSQSRGL 300
NTIALGTASN ATGDKSLALG SNSSANGINS VALGADSIAD LDNTVSVGNS SLKRKIVNVK 360
25 NGAIKSDSYD AINGSQLYAI SDSVAKRLGG GAAVDVDDGT VTAPTYNLKN GSKNNVGAAL 420
AVLDENTLQW DQTKGKYSA HGTSSPTASV ITDVADGTIS ASSKDAVNGS QLKATNDDVE 480
ANTANIATNT SNIATNTANI ATNTTNITNL TDSVGLDQAD ALLWNETKKA FSAAHGQDIT 540
SKITNVKADAD LTADSTDAVN GSQLKTTNDA VATNTTNIAN NTSNIATNTT NISNLTETVT 600
NLGEDALKWD KDNGVFTAAH GTETTSKITN VKDGLTTGS TDAVNGSQLK TTNDVATNT 660
30 TNIAATNTNI SNLTETVTNL GEDALKWDKD NGVFTAAHGN NTASKITNIL DGTVTATSSD 720
AINGSQLYDL SSNIATYFGG NASVNTDGVF TGPTYKIGET NYYNVGDALA AINSSFSTSL 780
GDALLWDATA GKFSAKHGTN GDASVITDVA DGEISDSSSD AVNGSQLHGV SSYVVDALGG 840
GAEVNADGTI TAPTYTIANA DYDNVGDALN AIDTTLDDAL LWDADAGENG AFSAAHGKDK 900
TASVITNVAN GAISAASSDA INGSQLYTTN KYIADALGGD AEVNADGTIT APTYTIANAE 960
35 YNNVGDALDA LDDNALLWDE TANGGAGAYN ASHDGKASII TNVANGSISE DSTDAVNGSQ 1020
LNATNMMIEQ NTQIINQLAG NTDATYIQEN GAGINYVRTN DDGLAFNDAS AQGVGATAIG 1080
YNSVAKGDSS VAIGQGSYSD VDTGIALGSS SVSSRVIAGK SRDTSITENG VVIGYDITDG 1140
ELLGALSIGD DGKYRQIINV ADGSEAHDAV TVRQLQNAIG AVATTPTKYF HANSTEEDSL 1200
AVGTDSLAMG AKTIVNGDKG IGIGYGAYVD ANALNGIAIG SNAQVIHVNS IAINGSTTTT 1260
40 RGAQNTYTAY NMDAPQNSVG EFSVGSADGQ RQITNVAAGS ADTDAVNVGQ LKVTDAQVSQ 1320
NTQSITNLDN RVTNLDSRVN NIENGIGDIV TTGSTKYFKT NTDGVDASAQ GKDSVAIGSG 1380
SIAAADNSVA LGTGSVATEE NTISVGSSTN QRRITNVAAG KNATDAVNVA QLKSSSEAGGV 1440
RYDTKADGSI DYSNITLGGG NGGTTRISNV SAGVNNNDVV NYAQLKQSVQ ETKQYTDQRM 1500
VEMDNKLSKT ESKLSGGIAS AMAMTGLPQA YTPGASMASI GGGTYNGESA VALGVSMVSA 1560
45 NGRWVYKLQG STNSQGEYSA ALGAGIQW 1588
<212> Type : PRT
<211> Length : 1588
SequenceName : SEQ ID 280
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 MPASAVGALG EASYTVTANV TDSAGNSNSA SHNVQVNTAL PGVTINPVAT DDIINAAESG 60
NAQTISGQVT GAAAGDTVTV TLGGKTYTAT VQGNLSWSVD VPAADIQAIG NGNLTVNASV 120
TNGVGNTGSG SRDITIDANL PGLRVDTVAG DDVNSIEHA QALVITGSSS GLAAGAALT 180
VINTVTYAAAT VLADGTWSVG VPAADVSNWP AGTVNITVSG TNTAGTTSTI THPVTVDLAA 240
VAISINTVSG DDVINAEEKG ADLTLGSGTS GVEVGQTVTV TFGGKTYTAT VAGDGSWTTT 300
60 VPAADLSVLR DGDA TVQASV STINGNTASA THAYSVDATA PTLAINTIAT DDILNAAEAG 360
NPLTISGSST AEAGQTVTVT LNGVTYSGSV QADGSWSVSL PTADLSNLTA SQYTVSASVS 420
DKAGNPASAN HGLAVDLTVP VLTINTVSGD DIINAAEHGQ ALVISGSSTG GEAGDVITVT 480
LNSKTYTTML DASGNWSVG VPAADV TALGS GPQTITAAIT DAAGNSDDAS RTVTVNLAAP 540
TIGINTIATD DVIKATEKGA DLQITGTSNQ PAGTTITVTL NGQNYTATTD SNGNWSATVP 600
65 ASAVSALGEA NYTVTANVTD TAGNSNSASH NVLVNSALPA VTINAVATDD IINAAESGNA 660
QTISGQVTGA AQGDTVTVTL GGNTYTATVQ SNLSWSVDVP AADIQALGNG DLTVNASVTN 720
GVGNTGSGSR DITIDANLPG LRVDTVAGDD VINSIEHNQA LVITGSSSGL TAGTALTVEI 780

| | | | | | | | |
|----|-------------|-------------|-------------|-------------|-------------|-------------|------|
| | NNVTYGATVL | ADGTWSLGVP | AVDVSNNWPAG | TVNITVSGTN | SAGTTSTITH | PVTVDLAGVA | 840 |
| | ITINTLSGDD | VINAVEKGET | LVVSGSTSGV | EAGQTVTVTF | GGKNYTTTVE | ANGSWTVNVP | 900 |
| | PADLAALPDG | AGNVQASVSN | INGNSAQADR | AYSVDATAPL | VTINTIASDD | ILNVSEAGAG | 960 |
| | ITISGTTTAAQ | AGQTLTVTLN | NNTYQTTVLA | DGTWSVNVPA | ADLSGLTASS | YTVTATVSDK | 1020 |
| 5 | AGNPASADHA | LVVDITAPDL | TINTVAGDDI | INAEHGGQAL | VVSGTSTGAA | AGDVVTVTLN | 1080 |
| | GKNYTTTTLDA | SGNWSVGIPA | ADVTALATGS | QTITASLSDR | AGNSDSTTHD | VTVDLSGPTL | 1140 |
| | TINTVSGDDI | INAAEIVVAQ | TISGQVTGTA | VAGNTVIVTI | GGNQYNATVQ | SDLWSVSVSP | 1200 |
| | ANVLQALGNG | ELTISASLTN | SANNTGTATH | DIVIDANLPG | LRVDTVAGDD | VINSIEHTQA | 1260 |
| | LVITGSSSGL | AAGAALTIVI | NSVTYGATVL | ADGSWSVGVP | VADVTNWPAG | TVNIAVSGTN | 1320 |
| 10 | TAGTTTTSISH | PVTVDLAAVA | ITINTLSTDD | VINAAEKGSD | LQLSGTTSVG | EAGQTITVIF | 1380 |
| | GGKSYTTTVA | ADNTWGLTIP | AVDVATLPDG | AANVQASVSN | VAGNSTQATH | AYSVDATAPS | 1440 |
| | VTINTIATDD | ILNAAEAGSA | LTISGTSTAE | AGQTVTVTLN | GVNYSGNVQA | DGSWSVSVPT | 1500 |
| | GDLASLTASS | YTVNASVSDK | ARNSASATHN | LTVDLAAPVV | TINTVAGDDI | INATEHGQAAQ | 1560 |
| | IISGSATGAT | TGNTVSVTIG | TTTTTTVLDA | NGNWSIGVPA | SVISALAAQGD | VTITATVTDS | 1620 |
| 15 | AGNSGTASHT | VTVALGAPVL | AINITIAVDDI | INAAEKGADL | AITGTSNQPA | GTQITVTLNG | 1680 |
| | QNYTTTADAS | GNWSVTPPAS | RVSALGEATY | TVTAAATDAD | GNSGSASHNV | QVNTALPGVT | 1740 |
| | INVVATDDII | NAAEAGVEQT | ISGQVTGAAA | GDTVTVTLGG | ATYTATVQAN | LSWSVDVPAS | 1800 |
| | ALQELGNGEL | TISASVTNSV | GNTGNGTREI | TIDANLPGLR | VDTVAGDDVV | NIEHGGQALV | 1860 |
| | ITGSSSGLAA | GSNVTLTING | QTYVAAVLAD | GTWSVGVPAPV | DVSAWPAGSV | TIAASGSTSA | 1920 |
| 20 | GNPVSVTHPV | TVDLASAVAVS | INAITADDDI | NAAEKGAAAL | LSGSTSGVEA | GQTVTVTFGG | 1980 |
| | KTYSATVAAN | GSWSTSVPA | DMAALRDGDA | SAQASVSNVN | GNSATTHAY | SVDASAPTVT | 2040 |
| | INTIAGDDIL | NAAEAGAALT | ITGSSTAEAG | QTVTVTLNGT | NYTGTVQTDG | SWSVSVPSAD | 2100 |
| | LSTLTASNYT | VNAAVSDKAG | NPASVNHNL | VDTSVPVVTI | NTVAGDDVIN | ATEHAQAQII | 2160 |
| | SGSATGAATG | STVTVTIGTN | TFTTVLDASG | NWSVGVPASV | VSALANGTVT | INASVTDAGG | 2220 |
| 25 | NSGSATHQVT | VNTGLPTITF | NAISGDNLIN | ADEKGQPLTI | SGGSTGLATG | AQVTVTLNGH | 2280 |
| | NYSATTDASG | NWTLTVPVSD | LAALGQANYT | VSASATSAAG | NTASSQANLL | VDSGLPDVTI | 2340 |
| | NTVAGDDIIN | AAEAGADQTI | SGVVTRAAAG | DTVTVTLGGN | TYTATVQSNL | SWSVSVPTAD | 2400 |
| | LQALGNGDLT | ITASVTNANG | NTGSGTRDIT | IDANLPGLRV | DTVAGDDIVN | SIEHGGQALVI | 2460 |
| | TGGSSGLNAG | AVLTVTINSV | AYSATVQADG | SWSVGIPAAN | VSAWPAGPLT | VEVDGQSSAN | 2520 |
| 30 | NPVSVSHPFT | VDLTAVAI | NTVASDDVIN | AAEKGNTLTL | SGSTSGIESG | QTVTVTFGGK | 2580 |
| | TYTASVAANG | SWSVNVPAAD | LATLPEGAAN | VQASVSSASG | NSASATHAYS | VDASAPTLTI | 2640 |
| | NTIASDDILN | AAEAGSPLTI | SGTSTAETGQ | TVTVTLNGAT | YTGTVQADGS | WSVSVPTSAL | 2700 |
| | GALNASNYTV | SATVNDKAGN | PGSASHNLAV | DTTAPVLTIN | TVAGDDIIND | AEHAQALVIS | 2760 |
| | GTSSGGEAGD | VVSUVLNGKT | YTTTLDASGN | WSVGVPAAADV | TALGSGAQTI | TASVSDRAGN | 2820 |
| 35 | SDDASRTVTV | SLSAPVISIN | TIAGDDVINA | TEKGSDDLALS | GTSDQPAGTA | ITVTLNGQNY | 2880 |
| | SATTDASGNW | SVTVPASAVS | ALGEATYSVT | ASVTNAQONS | STASHNVQVN | TALPGITINP | 2940 |
| | VATDDIINAS | EAGSAQTISG | QVTGAAAGST | VTVELGGKTY | TATVQADLSW | NVSVPAADWQ | 3000 |
| | ALGNGELTVN | ASVTNAVGN | GSSTRDITID | ASLPLGLRVD | VAGDDVNI | EHAQAQVITG | 3060 |
| | SSSGFAAGTA | LTVVINNQTY | AATVLANGSW | SVGVPAATDVS | NWPAGTLNIT | VSGANSAGTQ | 3120 |
| 40 | TSITHPLTVD | LTAVAI | ITSDDAINAA | EKGAAALTLG | STSGVEAGQT | VTVTFGGKTY | 3180 |
| | TTTVAANGSW | STTVPAADLA | ALRDGDASAQ | VRVTNVNGNS | ATATHEYSVD | SAAPTVTINT | 3240 |
| | IASDNIINAS | EAAAGVTVSG | TSTAQTGQTL | TVTLNGTNYQ | TTVQTDGWS | LTLPASDLTA | 3300 |
| | LANNGYTLTA | TVSDLAGNLG | SASKGVTVDT | TAPVISFNTV | AGDDVINNV | HIQAQIISGT | 3360 |
| | ATGAVAGDRL | VVTIAGQQYV | TSTDASGNWS | VGVPASVISG | LADGTVTISA | TITDSAGNSS | 3420 |
| 45 | TQTHNVQVNT | AAVSLSVSTI | SGDNLINAAE | AGSALTLSGT | GTNFATGTVV | TVLLNGKGYS | 3480 |
| | ATIQSNWSWS | VNVPAADVAA | LSDGTSTYVS | ASAQDSAGNG | NSSTQTHNVQ | VNTAAVSLSV | 3540 |
| | STISGDNLIN | AAEAGSALT | SGTGTNFATG | TVVTVLLNGK | GYSATIQSNG | SWSVNVPAAD | 3600 |
| | VAALSDGTSY | TVSASAQDSA | GNSATASRSV | AVDLTAPVIS | INTVSTDDRL | NAAEQQQPLT | 3660 |
| | LNGSTSAEVG | QTVTVTFGGK | TYTATVAANG | TWALNVPAVD | LAALGQGAQT | ITASVNDRAG | 3720 |
| 50 | NPGQATHALT | VDTVAPTITI | ATVAGDDIIN | NAEQLAQQTI | SGTTTAEVQ | TVTTVTFNGQT | 3780 |
| | WSATVSGSGS | WSVFIPAQQF | AGLSDGSYTI | SATVSDQAGN | PGSASRGVTL | NGDVPTVTIN | 3840 |
| | TFAGDDVUNA | AEHGSSLVIS | GTTPAPVGQT | LTLTLNGKTY | TTTVQTGGSW | SYTLGSADVT | 3900 |
| | ALADGNAYVI | NASVSNAIGN | TGSSNHTITV | DLSAPAMGIN | IDSLQADTGL | SASDFITSVS | 3960 |
| | PVVVNGSLTA | ALASNETAQI | SIDGGTTWTT | LTVTGTTWRY | NDSRTLTDGN | YLYQVRVIDA | 4020 |
| 55 | AGNVGATDSQ | NVIDTTAPD | PAVKTIAISA | ITDMGLITN | DFVTSDDTLA | VSGTLGATLS | 4080 |
| | AGEFAQISLD | GGVTWTTLT | VGTWSYADG | HTLTDGTWNY | TVRVVDLAGN | VGQTATQNVV | 4140 |
| | VDTTSPAAK | SITITGISDD | TGTSSSDFIT | SDTTLTVRGV | LGAALGANEF | AQISTDNGAT | 4200 |
| | WVNVTVAA | LNWSYVDGRT | LTNGTTTWQV | RVVDLAGNVG | ATSSQSALID | TVNPAQVLT | 4260 |
| | ASISTDTGSS | ATDFITSMT | LTLTGSLGAG | LASGEVAQIS | LDGATWTTT | TNGTQWTYT | 4320 |
| 60 | DSRTLTDGSY | VYQVRVLDLA | GNTGPVVSKT | VVVDTINPTA | TPTIVSYTDD | VGQRQGTLS | 4380 |
| | SQATDDTTPL | LNGVLSAPLA | SGEVVLYLRY | GLLLGAVTMV | GALNWTYS | GLVSGAYTYS | 4440 |
| | ARVVDLAGNI | TSSSDFVLTV | DTSIPTTLAQ | ITSQTTTRDT | PIISGVITAA | LASGQYVEV | 4500 |
| | INGKTYTSEP | GGAVVVDPAH | NTWYVQLPDT | DALTVSATA | TVTAQVKSSA | GNGNNANIS | 4560 |
| | GTVTVNAAID | YTPTWTTASK | TTAWGLTYGL | DSHGMWTVLA | NQQVMQSTDP | LTWSKTALTL | 4620 |
| 65 | YQSGNNYATS | SIADYDRNGT | GDLFITRDDY | GTGYINGFTN | NGDGTFS | QVTVGTLTWY | 4680 |
| | GSIVAFDKEG | DGYLDFWIGD | AGGPDSNTFL | WNNAGTLVGN | STTSNSGGSA | TVGGAVTGYL | 4740 |
| | SLNEGSGVDL | NNDGRIDLQ | HTYNLNNYYT | LSSLINQNG | TFVWGQNTTN | TFLSGAGSGA | 4800 |

MSSSVSMTWA DFDGDGDMDL FLPASQGRAN YGSLLFNTNG VLGCPVAVGA TATTYASQFS 4860
LAVDWNHDGL MDIARIAQTG QSYLYTNVSN ASNWTQSALG GSQSGTTSKV AAMDYDWDGA 4920
VDVLVSKQSG SVFLSRNTNT VSYGTSLSHLR ITDPNGINNVY YGNTVKLYNS AGVLVATQII 4980
NPQSGMGVND TSALVNFYGL NAGETYNAVL IKSTGTTASN IDQTVNTSWG GLQATDATHA 5040
5 YDLSAEAGTA SNNGKFVGTG YNDTFFATAG TDTYDGSQGW VYSSGTGTWL ANGGMDVVDF 5100
RLSTVGV TAN LSSTAAQATG FNTSTFTNIE GISGSNFNDI LTGSSGDNQL EGRGGNDTLN 5160
IGNGGHDTLL YKLLNASDAT GGNGSDVVNG FTVGTWEGTA DTDRIDIREL LQSGGYTGNG 5220
KASYVNGVAT LDAQAGNIGD FVKVTQSGSD TIVQIDRDGT GGTTFATTNVV TLTGVHTDLA 5280
TLLANHQLMV V 5291
10 <212> Type : PRT
<211> Length : 5291
SequenceName : SEQ ID 281
SequenceDescription :
15 Sequence

<213> OrganismName : Escherichia coli 0157:H7
<400> PreSequenceString :
20 MGVHTAEATL PNGNNDTKIV NIAPDASNAQ VTLNIPAAQV VTNNSDSVQL TATVKDPSNH 60
PVAGITVNFT MPQDVAANFT LENNGIAITQ ANGEAHVTLK GKKAGTHTVT ATLSNNNTSD 120
SQPVTFVADK TSALVVLQIS KNEITGNGVD SATLTATVKD QFDNEVNNLP VTFSTASSGL 180
TLTPGESNTN ESGIAQATLA GVAFGQTVT ASLANNGASD NKTVHFIGDT AAKIIEITP 240
VPDSIIAGTP QNSSGVSITA TVVDNNGFPV KGVTVNFTSN AATAEMTNGG QAVTNEQGKA 300
TVTYTNTRSS IESGARPDTV EASLENGSST LSTSINVNAD ASTAHLTLQ ALFDTVSAGD 360
25 TTNLYIEVKD NYGNGVPQQE VTLSPVSPSEG VTPSNNAIYT TNHDGNFYAS FTATKAGVYQ 420
VTATLENGDS MQQTVTYVPN VANAELSLAA SKDPVIANNN DLTTLTATVA DTEGNAIANS 480
EVTFTLPEDV RANFTLGDGG KVVTDTEGKA KVTLLKGTAG AHTVTASMAK GKSEQLVVNF 540
IADTLTAQVN LNVTEDNFIA NNVGMTRLQA TVTDGNGNPL ANEAVTFTLP ADVSASFTLG 600
QGGSAITDIN GKAEVTLSTG KSGTYPVTVS VNNYGVSDTK QVTLIADAGT AKLASLTSVY 660
30 SFVVSTTEGA TMTASVTDAN GNPVEGIKVN FRGTSVTLSS TSVETDDRGE AEILVTSTEV 720
GLKTVSASLA DKPTEVISRL LNAKADINSA TITSLEIPEG QVMVAQDVAV KAHVNDQFGN 780
PILNESVTFS AEPPEHMTIS QNIVSTDTHG IAEVTMTPER NGSYMKASL ANGSSYEKDL 840
VVIDQKLTLA ASSPLIGVNS PTGATLTATL TSANGTPVEG QVINFSVTPE GATLSGGKVR 900
TNSSGQAPVV LTSNKVGYT VTASFHNGVT IQTQITVKVT GNSSTAHVAS FIADPSTIAA 960
35 TNSDLSTLKA TVEDGSGNLI EGLTVYFALK SGSATLSTLT AVTDQNGIAT TSVRGAITGS 1020
VTVSAVTTAG GMQTVDTITLV AGPADASQSV LKNNRSSLKG DFTDSAEHL VLHDISGNPI 1080
KVSEGLEFVQ SGTNAPYVQV SAIDYSKNFS GEYKATVTGG GEGIATLIPV LNGVHQAGLS 1140
TTIQFTRAED KIMSGITLVN GANLPTTTTFP SQGFTGAYYQ LNNDNFAPGK TAADYEFSSS 1200
ASWVDVDTAG KVTFTKNGSK WERITATPKT GGPSYIYEIR VKSWVWVAGD AFMIYSLAEN 1260
40 FCSSNGYTLP LGDHLNHSRS RGIGSLYSEW GDMGHYTTEA GFHSNMYWSS SPANSNEQYV 1320
VSLATGDQSV FEKLGFAIAT CYKNL 1345
<212> Type : PRT
<211> Length : 1345
SequenceName : SEQ ID 282
45 SequenceDescription :
Sequence

<213> OrganismName : Escherichia coli 0157:H7
50 <400> PreSequenceString :
MSLIIDVISR KTSVKQTLIN PGDVTVVIYE PSVVQVHAQA SAVARYVREG NDLLIYMQDG 60
TVIRCNGYFL QAANTAEQSE LVFADGQQLT HITFADTAAG GLAPVELTAQ TTAIESIAPF 120
LDTVAQTSFA PWGWLAAAV GGGALGALLA SGGDGDSTKE VINNPTPPAE PGNATPSFLV 180
TDNQGDQIRI LATNDITDDT TPTFSGSQQA GATIIQKDSN GNTIASTQVD NNGHWSVSLP 240
55 TQSAGEHTWS VVQIVGSTIT DAGSITLTID NSQASVQVAT TAGDNIINAS EQAAGFTLSG 300
TSSHLAQGE LTVTLNGKTY TTSVGANGAW SVQVPTADAQ ALGEGNQAVL VSGKDATGNT 360
VTGAQLLTVD TQPPTLAINT IAQDNIIISAA EHNVALVLSG TSNAEAGQTV TLTVNGKSHT 420
ATVGSDDGTWQ VTLPADEVQA LAEGNYAVNA SVSDRAGNTT SHSANFTVDT SAPVVSNTV 480
AGDDILNNAE QAVAQIISGQ VSGASPGDTV TVKLGTHTLT GIVLADGSWN VALDPAVTRT 540
60 LDRGANTIFV TVTDAAGNTG AASRAITLVG VSPLITINTV SGDDIISGAE KGAPLTLTGS 600
TQQAETGQTV TVTLAQSFST TTVQADGSWS LTVPAAMGN LPDGAVAITA SVTDLSGNTG 660
NTSRTITVDS QAPALSIDPL TADNIINAAE SGQDLPTGT TDAQPGQTV TLTNGQTYQG 720
VVQPDGTWSV TVPAANVGAL ADGNATVTAS VNDVAGNPSS VSRVALVDAT PPVVTINPVA 780
TDNVINTPEH AQAQIISGTV TGAQAGDIVT VTLNNDVYTT VVDGSGNWSL GVPASVVSGL 840
65 ADGSYPVSVS VTDKAGNTGS QSLTIVTNTA APLIGINSIA GDDVINASEK GADLQITGTS 900
DQPVNTAITV TLNGQNYTTT TDASGNWSVT VPASAVTALG QANYTVTAAV TSDIGNSATA 960
SHNVLVDSAL PGVTINPVAT DDIINAAEAG VAQTISGQVT GAEDGDTVIT TLGGNTYTAT 1020

VGSNLTWSVD VPAADIQALG NGDLTVNASV TNQNGNTGSG TRDITIDANL PGLRVDTVAG 1080
DDVNNIEHG QALVVTGSSS GLAESTPLTV TINNVEYTTA VQADGSWSVG VTAAQVSAWP 1140
AGTVNIAVSG ESSAGNSVSI THPVTVDLTP AAITINTIAT DDVINAEEKG ADLTLSGTTT 1200
NVEPGQTVTV TFGGKNYTAS VASDGSWTAT VPAADLASLP EGSASALASV SNINGNSASA 1260
5 VHNYSVDSSA PTIIINTVAS DNIVNASEAD AGVTVSGSTT AEAGQIVTIT LNSPTVQTYQ 1320
ATVQADGSWS INIPAADLEA LTDGSHTLTA TVNDKAGNPA STHNLAVDL TVPVLTTINTI 1380
AGDDIINATE HGQALVISGS STGGEAGDVV TVTLNSKTYT TTLDASGNWS VGVPAADVTA 1440
LGSGPQTVTA TVTDAAGNSD N 1461
<212> Type : PRT
10 <211> Length : 1461
SequenceName : SEQ ID 283
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNRIYRVIWN CTLQVFQACS ELTRRVGKTS TVNLRKSSGL TTKFSRLTLG VLLALSGSVS 60
GASLEVDNGQ ITNIDTDVAY DAYLVGWYGT GVLNILAGGN ASLTTITTSV IGGNEDSEGT 120
20 VNVLGGTWRL YDSGNNARPL NVGQSGTGTL NIKQKGHVDG GYLRLGTQAA GVGTVNVEGE 180
DSVLTTELF EIGSYGTGSLN ITDKGYVTSS IVAILGYQAN SNGKVVEKG GEWLIKNNDS 240
SIEFQIGNQG TGEATIREGG LITAENTIIG GNATGVGTLN VQDQDSVITV RRLYNGYFGN 300
GAVNISNNGL INNKEYSLVG VQDGSBGVVN VTDKGHWNFL GTGEAFRIY IGDAGDGELN 360
VSREGKVD SG IITAGMKETG TGNLTVKDKN SVITNLGTNL GYDGHGEMNI SNEGLVVSNG 420
25 GSSLGYGETG VGKVSITTGG IWEVNKNVYT TIGVAGVGNL NISDGGKFVS QNITFLGDKA 480
SGIGTLNLM D ATSSFDTVGI NVGNFGSGIV NVSNGATLNS TGYGFIGGNA SGKGINVIST 540
DSLWNLKTSS TNAQLLQGVV LGTGELNITT GGIVKARDTQ IALNDKSKGD VRVDGQNSLL 600
ETFNMYVGT S GTGTLTLTNS GTLNVEGGEV YLGVFEPAVG TLNIGAAHGE AAADAGFITN 660
ATKVEFGSGE GVFVFNHTNN SDAGYQVDM L ITGDDKDGV IHDAGHTVFN AGNTYSGKTL 720
30 VNDGLLTIAS HTADGVTGMG SSEVTIASPG TLDILASTNS AGDYTLTNAL KGDGLMRVQL 780
SSSDKMFEGT HATGTEFAGV AQLKSTFTL ERDNTAALH AMLQSDIENT TSVNVGEQSI 840
GGLAMNGGTL IFDTPAAT LAEGYISVD L LVVGASDYTE KGRNYQVNGT GDVLIGVPKP 900
WNDPMANNPL TTLNLLEHDD NHVGVQLVKA QTVIGSGGSL TLRDLQDEV EADKTLHIAQ 960
NGTVVAEGDY GFRLTAPGD GLYVNYGLKA LNIHGGQKLT LAEHGGAYGA TADMSAKIGG 1020
35 EGDLAINTVR QVSLNGQND YQGATYVQMG TLRTDADGAL GNTRELNISN AAIVDLNGST 1080
QTVETFTGQM GSTVLFKEGS LTVNKGGISQ GELTGGGNLN VTGGTLAVEG LNARYNALTS 1140
VSPNAEVSLD NTQGLGRGNI ANDGLLTLKN VTGELRNSIS GKGIVSATAR TDVELDGDNS 1200
RFVQGQFNIDT GSALSVNEQK NLGDASVINN GLLTISTERS WAMTHSISGS GDLTKLGTGI 1260
LTLNNDSSAY QGTTDIVGGE IAFGSDSAIN TASQHINIHN SGVMSGNVTT AGDVNVMSGG 1320
40 TLRVAKTTIG ESAATWRMAA RFK 1343
<212> Type : PRT
<211> Length : 1343
SequenceName : SEQ ID 284
SequenceDescription :

Sequence
45 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 MGIKQHNGNT KADRLAELKI RSPSIQLIKF GAIGLNAILF SPLLIAADTG SQYGTNITIN 60
DGDRITGDTA DPSGNLYGVM TPAGNTPGNI NLGNDVTVNV NDASGYAKGI IIQGNSSLT 120
ANRLTVDVVG QTSAGINLI GDYTHADLGT GSTIKSNDG IIIGHSSTLT ATQFTIENS 180
GIGLTINDYG TSVDLGSGSK IKTDGSTGVY IGGLNGNNAN GAARFTATDL TIDVQGY SAM 240
GINVQKNSV DLGTNSSIKT SGD NAHGLWS FGQVSANALT VDV TGAANG VEV RGTTTI 300
55 GADSHISSAQ GGGLVTSGSD ATINFSGTAA QRNSIFSGGS YGASAQTATA VINMQNTDIT 360
VDRNGSLALG LWALSGGRIT GDSLAITGAA GARGIYAMTN SQIDLTS DLV IDMSTPDQMA 420
IATQHDDGYA ASRINASGRM LINGSVLSKG GLINLDMHPG SVWTGSSLSD NVNGGKLDVA 480
MNNVWNVTS NSNLDLALS HSTVDFASHG STAGTFTTLN VENLSGNSTF IMRADVVGE 540
NGVKPWA 547
60 <212> Type : PRT
<211> Length : 547
SequenceName : SEQ ID 285
SequenceDescription :

Sequence
65 -----
<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :
MGIDSRNDIP EGIATLGAFM GYSHSHIGFD RGGHGSVDSY SLGGYASWEH ESGFYLDGVV 60
KLNRFESNVA GKMSSGGAAN GSYHSNGLGG HIETGMRFTD GNWNLTPTYAS LTGFTADNPE 120
YHLSNGMESK SVDTRSIYRE LGATLSYNMR LGNGMEVEPW LKAAVRKEFV DDNRVKVNSD 180
5 GNFDVNDLSGR RGIYQAGIKA SFSSTLSGHL GVGYSNGAGM ESPWNAVAGV NWSF 234

<212> Type : PRT
<211> Length : 234
SequenceName : SEQ ID 286
10 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
15 <400> PreSequenceString :
MKKKVLAIAL VTVFTGMGVA QAADVTAQAV ATWSATAKKD TTSKLVVTPL GSLAFQYAEG 60
IKGFNSQKGL FDVAIEGDST ATAFKLTSRL ITNTLTQLDT SGSTLNVGVD YNGAAVEKTG 120
DTVMIDTANG VLGGNLSPLA NGYNASNRRT AQDGFTFSII SGTNGTTAV TDYSTLPEGI 180
WSGDVSVQFD ATWTS 195
20 <212> Type : PRT
<211> Length : 195
SequenceName : SEQ ID 287
SequenceDescription :

25 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MTAESYDDNY LDDEDADWTA TGQGQKSAGD TSFTLAWKPG EEGQKGLIGW FESGDVRAYK 60
30 IRFPNGTVDV FRGWVSSIGK AVTAKEVITR TVKVTNVGKP SVAEERSKIT PVSAIKVTPT 120
SGTVAKGKTT TLTVSFEPES ATDKTFRAVS ADPSKATISV KDMTITVNGV ATGKVQIPVV 180
SGNGQFAAVA EVTVTEAGAA G 201
<212> Type : PRT
<211> Length : 201
35 SequenceName : SEQ ID 288
SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MTAESYDDNY LDDEDADWTA TGQGQKSAGD TSFTLAWKPG EEGQKGLIGW FESGDVRAYK 60
IRFPNGTVDV FRGWVSSIGK AVTAKEVITR TVKVTNVGKP SVAEERSKIT PVSAIKVTPT 120
SGTVAKGKTT TLTVSFEPES ATDKTFRAVS ADPSKATISV KDMTITVNGV ATGKVQIPVV 180
45 SGNGQFAAVA EVTVTEAGAA G 201
<212> Type : PRT
<211> Length : 201
SequenceName : SEQ ID 289
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 MLYNIPCRIY ILSTLSLCIS GIVSTATATS SETKISNEET LVVTNRSAS NLWESPATIQ 60
VIDQQTQLQNS TNASIADNLQ DIPGVEITDN SLAGRKQIRI RGEASSRVL I LIDGQEVTYQ 120
RAGDNYGVGL LIDESALERV EVVKGYPYSL YGSAIGGIV NFITKKGGDK LASGVVKAVY 180
NSATAGWEES IAVQGSIGGF DYRINGSYSD QGNRDTDPGR LPNTNYRNNS QGVWLGYNSS 240
NHRFGLSLDR YRLATQTYE DPDGSYEAFS VKIPKLEREK VGVFYD TDVD GDYLKKIHFD 300
60 AYEQTIQRQF ANEVKTTQPV PSPMIQALTV HNKTDTHDKQ YTQAVTLQSH FSLPANNELV 360
TGAQYKQDRV SQRSGGMTSS KSLTGFINKE TRTRSYYESE QSTVSLFAQN DWQFADHWTW 420
TMGVRQYWLS SKLTRGDGVS YTAGIISDTS LARESASDHE MVTSTSLRYS GFDNLELRAA 480
FAQGYVFPTL SQLFMQTSAG GSVTYGNPDL KAEHSNNFEL GARYNGNQWL IDSAVYYSEA 540
KDYIASLICD GSIVCNGNTN SSRSSYYYD NIDRAKTWGL EISAEYNGWV FSPYISGNLI 600
65 RRQYETSTLK TTNTGEPAIN GRIGLKHTLV MGQANIISDV FIRAASSAKD DSNGTETNVP 660
GWATLNFVN TEFGNEDQYR INLALNNLTD KRYRTAHETI PAAGFNAAIG FVWNF 715

<212> Type : PRT
<211> Length : 715
SequenceName : SEQ ID 290
SequenceDescription :

5
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
10 MTKMSRYALI TALAMFLAGC VGQREPAPVE EVKPAPEQPA EPQQPVPTVP SVPTIPQQPG 60
PIEHEDQTAP PAPHIRHYDW NGAMQPMVSK MLGADGVTAG SVLLVDVNN RTNGSLNAAE 120
ATETLRNALA NNGKFTLVSA QQLSMAKQQL GLSPQDSLGT RSKAIGIARN VGAHYVLYSS 180
ASGNVNAPTL QMQLMLVQTG EIIWSGKGAV SQQ 213
<212> Type : PRT
15 <211> Length : 213
SequenceName : SEQ ID 291
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
20 MKSKVLALLI PALLGAGAAH AAEVYNKDN KLDLYGKVDG LHYFSDNSAK DGDQSYARLG 60
FKGETQINDQ LTGYGQWEYN IQANNTESK NQSWTRLAFA GLKFSDYGSF DYGRNYGLDR 120
25 YAA 123
<212> Type : PRT
<211> Length : 123
SequenceName : SEQ ID 292
SequenceDescription :

30
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 MATPNPLEPV KGAGTTLWVY NGKGDAYANP LSDDDWQRLA KVKDLTPGEM TAEPYDDNYL 60
DDEDADWTAT GQGQKSAGDT SFTLAWKPGE EGQKGLIGWF ESGDVRAKY RFPNGTVDVF 120
RGWVSSIGKA VTAKEVITRT VKVTNVGKPS VAEERSEITP ATAIVTPTS GTVAKGKTTT 180
LTVSFEPESA TDKTFRAVSA DPSKATISVK DMTITVNGVA TGKVQIPVVS GNGQFAAVAE 240
VTVTEAGAAG 250
40 <212> Type : PRT
<211> Length : 250
SequenceName : SEQ ID 293
SequenceDescription :

45
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 MATPNPLEPV KGAGTTLWVY NGKGDAYANP LSDDDWQRLA KVKDLTPGEM TAEPYDDNYL 60
DDEDADWTAT GQGQKSAGDT SFTLAWKPGE EGQKGLIGWF ESGDVRAKY RFPNGTVDVF 120
RGWVSSIGKA VTAKEVITRT VKVTNVGKPS VAEERSEITP ATAIVTPTS GTVAKGKTTT 180
LTVSFEPESA TDKTFRAVSA DPSKATISVK DMTITVNGVA TGKVQIPVVS GNGQFAAVAE 240
VTVTEAGAAG 250
<212> Type : PRT
55 <211> Length : 250
SequenceName : SEQ ID 294
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
60 MGWTDMLPEF GGDSYTNADN FMTGRANGVA TYRNTDFFGL VNGLNFAVQY QGNNEGASNG 60
QEGTNNGRDV RHENGDWGL STTYDLGMGF SAGAAYTSSD RTNDQVNHTA AGGDKADAWT 120
65 AGLKYDANNI YLATMYSETR NMTPFGSDY AVANKTQNF VTAQYQFDFG LRPVSVFLMS 180
KGRDLHAAGG ADNPAGVDDK DLVKYADVGA TYYFNKNMST YVDYKINLLD EDDSFYAANG 240
ISTDDIVALG LVIYQF 255

<212> Type : PRT
<211> Length : 255
SequenceName : SEQ ID 295
SequenceDescription :

5
Sequence

<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :

10 MGFIMKLTKT ALCTALFATF TFSANAQTYP DLPVGIKGGT GALIGDTVYV GLGSGGDKFY 60
TLDLKDPSAQ WKEIATFPGG ERNQPVAAGV DGKLYVFGGL QKNEKGELQL VNDAYRYNPS 120
DNTWMKLPTR SPRGLVGSSG ASHGDKVYIL GGSNLSIFNG FFQDTVAAGE DKAKKDEIAA 180
AYFDQRPEDY FFTTELLSYE PSTNKWRNEG RIPFSGRAGA AFTIQGNELV VVNGEIKPGL 240
RTAETHQGKF TAKGVQWKNL PDLAPKPKGS QDGLAGALSG YSNGHYLVGT GANFPKSIKQ 300
15 FKEGKLHAHK GLSKAWHNEV YTLNNGKWRI VGELPMNIGY GFSVSYNNKV LLIGGETDGG 360
KALTSVKAIS YDGKKLTIE 379
<212> Type : PRT
<211> Length : 379
SequenceName : SEQ ID 296
SequenceDescription :

20
Sequence

<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :

25 MGEQYMLTTI LSFLIVTTVV AYVSWLKTG DDLKSSKGYF LAGRGLSGLV IGCSMVLTS 60
STEQ LIGVNA VSYKGNFSVI AWTVP TVIPL CFLALYIIGW L 101
<212> Type : PRT
<211> Length : 101
SequenceName : SEQ ID 297
SequenceDescription :

30
Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :

35 MKNQHKNP LT KALMKTPYN HFLFFCFILG AFLLG LLSPA YALSIITTKE IDANLLNGAI 60
ESRVVLGKRV FKVEAHGFYF RNNATNSIDI EITSLLRDNQ SFPLTSSAKT SLKIPPNAKI 120
KKSTILVLKG ENAEVAKIL GVSKEEYQKL ENIAQTAAAN DPMYANTPFS NGSDSSFYDN 180
40 NPNSPSNNAI NGKDGANGSN GYGANGNDGV NGISGSNGAN GSHSNNAIG SGIDTDGVLG 240
VDGVNGSSSS SGGSVGGYEN NFTNHGSTNN NTGGYDNFNN GSSSGGSLGN GGLFPIPGN 300
GDTNNSNNST NTTSPINGSS SNNATNPSSQ ENNYSSQYCK VPELSPNNTM KLDVIAKDG 360
CISMNALRDD TKCAYRYDFE AGKAIKQTQY YYVDRENKTQ NIGGCVDLQG AQYAMQLYKD 420
DSKCALQTTS DKG YGMGKTQ TFQTEIVFRG MDNLIHVA VP CSDYARVQDR IVRYEKNDKT 480
45 QTLTPIVDQY YNDPNNPNKQ EILNRGIATQ LSSQYQEFAC GQWEYND AKL EAKRPTMLKS 540
YNKLNGEWE VTPCNFEAGI KSGAVVSPYV MGVPSSKVL S DITTSHYFRI ERKNYGEREQ 600
CQKLYGVNRC QPQYSILILV SPIGAPLTKP LPPKPLNLIY AQP KIMKNTP QPIILSPLKP 660
PSTGLKAF 668
<212> Type : PRT
<211> Length : 668
SequenceName : SEQ ID 298
SequenceDescription :

50
Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :

55 MPVIRVLV ML ATMMMKLVKT AKEKKVFKNV GISIMGIAFW EAIKDSIKKQ IKKSDWICGN 60
VKTADDYLKT HPNSWFNSAI GVTAITAMLM NVCFADDQSK KEVAQAQKEA ENARDRANKS 120
60 GIELEQEEQK TEQEKQKTEQ EKQKTEQEKQ KTEQEKQKTE QEKQKTSNIE TNNQIKVEQE 180
QQKTEQEKQK TNNTQKDLVN KAEQNCQENH NQFFIKKLG I KAGIAIEIEA ECKTPKPTKT 240
NQTPIQPKHL PNSKQPHSQR GSKAQELIAY LQKELES LPY SQKAI AKQVD FYRPSSIAYL 300
ELDPRDFNAT EEWQKENLKI RSKAQAKMLE MRSLKPD PQA HLSTSQSLLL VQKIFADVSK 360
EIKVVAN TEK KVEKAGYGY S KRM 383
<212> Type : PRT
<211> Length : 383
SequenceName : SEQ ID 299

65

SequenceDescription :

Sequence

5 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MNYPNLPNSA LEISEQPEVK EITNELLKQL QNALRSNAHF SEQVELSLKC IVRILEVLLS 60
LDFFKNANEI DSSLRNSIEW LTNAGESLKL KMKEYERFFS EFNTSMHANE QEVNTNLNAN 120
AENIKSEIKK LENQLIETTT RLLTSYQIFL NQARDNANNQ ITKNKTQSLE AITQAKNNAN 180
10 NEISNNQTQA ITNITEAKTN ANNEISNNQT QAITNINEAK ESATTQINAN KQEAINNITQ 240
EKTQATSEIT EAKKTDHYQN IDFFEFE 267
<212> Type : PRT
<211> Length : 267
SequenceName : SEQ ID 300
15 SequenceDescription :

Sequence

20 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKFFSKDLFK KVTPLFLSVY FLSPTLTQAK SRFYVASQYQ VGKMIMKKYN DLKRTIEGAS 60
FSLGWEINPT NYWFYSRYYP FMDYGNVILN KRTGAQANMF TYGFGGDLIM EYNKNPLYVF 120
SLFYGMQVAE NTWTISKHSA NFIIDDWRSI QGFSLKTSNF RMLGLVGFKF QTVLFHHDAS 180
IEVGIKWPFA FEYDSPFVRL FSVFISHTFY L 211
25 <212> Type : PRT
<211> Length : 211
SequenceName : SEQ ID 301
SequenceDescription :

30 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKFTLSLFL CCTLLNAEED IFRNNTNETD LTNSFEHGKE NNNLIPAKSD SLESFKEQEN 60
35 KEKAKQLMDL KALQSVYFSK NRKLQDNNFN VLYVAGNTNK IRLRYAMTTT FIFDNDPIIY 120
VSLGDPDSDFE LTYPTNDHYD LSNMLVIKPL LIGVDTNLTV VGASGTIYTL LFV 173
<212> Type : PRT
<211> Length : 173
40 SequenceName : SEQ ID 302
SequenceDescription :

Sequence

45 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLDYVPWIGN GYRYGNNHRG SNSSTSGVTT QGQSQNASSN EPAPTFSNVG VGLKANVNGT 60
LSGSRTTPNQ QGTPWLTLDQ ANLQLWTGAG WRNDKNGQSD ENYTNFASAK GSTNQQGSTT 120
GGSAGNPDSL KQDKADKSGD SVTVAEATSG DNLNTYTNLP PTSPPHPTDR TRCHSPTRTT 180
50 PSGCSCSCAA CWAASRCWSI RVGKMITVSL IPPTKNGLTP N 221
<212> Type : PRT
<211> Length : 221
SequenceName : SEQ ID 303
SequenceDescription :

55

Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
60 MDDITAPQTS AGSSSGTSTN TSGSRSFLPT FSNVGVGLKA NVQGTLGGRQ TTTTGNNIPK 60
WATLDQANLQ LWTGAGWRND KTTSGSTGNA NDTKFTSATG SGSGQGSSSG TNSAGNPDG 120
LQADKVDQNG QVKTSVQEAT SGDNLNTYTN LPPANLTPTA DWPNALSFNT KNAQRAQLF 180
LRGLLGSIPV LVNKSQDDN SKFKAEDQKW SYTDLQSDQT KLNLPAYGEV NGLLNPALVE 240
TYFGNTRASG SGSNTTSSPG IGFKIPEQSG TNTTSKAVLI TPGLAWTPQD VGNIVVSGTS 300
65 FSFQLGGWLIV TFTDFIKPRA GYLGLQLTGL DVSEATQREL IWAKRPWAAF RGSWVNRLGR 360
VESVWDFKGV WADQAQLAAQ AATSSTTTTA TGATLPEHPN ALAYQISYTD KDSYKASTQG 420
SGQTNSQNNNS PYLHFIKPKK VESTTQLDQG LKNLLDPNQV RTKLRQSFGT DHSTQPPQPS 480

LKTTTPVFGR SSGNLSSVFS GGGAGGGSSG SGQSGVDLSP VERVSGH 527
<212> Type : PRT
<211> Length : 527
SequenceName : SEQ ID 304
SequenceDescription :

5 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
10 MLKLAVGIFI SPTLTRFSTG FNLAGSVLDQ VLDYVPWIGN GHRYGNNHRG VDDITAPKTG 60
AGSSSGTSTN TSGSRFLPT FSNVGVGLKA NVQGTLLGSSQ TTTTGKDIPK WPTLDPANLQ 120
LWTGAGWRND KASNQSDEN HTTFKSATGS GQGGSTTGG SAGNPDSLKQ DKISKSGQNL 180
TTQDGAPQSN STTESASNYD HLPNLTPTS DWPNALSTFN KNNAAQRAQLF LRGLLGSIPV 240
15 LVNRSGSDDS NKFQATDQKW SYTDLKSDQT KLNLPAYGEV NGLLNPALVE TYFGTTRAGG 300
SGSNTTSSPG IGFKIPEQNN DSKAVLITPG LAWTPQDVGN LVSSTSLSF QLGGLVLTFT 360
DFVKPRAGYL GLQLTGLDAS DATQRALIWA KRPWAAFRGS WVNRLGRVES VWDLKGWQD 420
QAQAAAQAAT TAAATGDALP EHPNALAYQI SSTDKDSYKA STQSSGQTNQ QNTSPYLHLI 480
KPKKVENTTQ LDQGLKTCWT PTRFAPSCAK ALVQTIPPKP NPNPSKQPHR CLGRIVVTLA 540
20 VCLVVGVEE QTAPIRWTSP PLNGWVGGLW GNYPVGVGGI VVRILKVCKT LFLISIFISI 600
FFLNCSLTLE IWTASLATG LTVVGHTST TTTLKRQQFS YTRPDEVALR HTNAINPRLT 660
PWTYRNTSFS SLPLTGENPG AWALVRDNTA KGITAGSGSQ QTTYDPTRTE AALTTATTFV 720
LRRYDLAGRC TTSTFRS 737
<212> Type : PRT
<211> Length : 737
SequenceName : SEQ ID 305
SequenceDescription :

30 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
35 MLDYIPWIGN GHRYGNDHRG SNSSTSGVTT QGQSQNASG TEPASTFSNV GVGLKANVQG 60
TLGGSQTTTT GKDIPKWPTL DQANLQLWTG AGWRNDKASS QSDENHTKF TSATGSGQQG 120
SSSGTTNSAG NPDLSKQDKV DKSQDSVTVA ETTSGDNLTN YTNLPPNLTP TADWPNALSF 180
TNKNNAQRAQ LFLRALLGSI PVLVNKSGQD DSNKFQATDQ KWSYTELKSD QTKLNLPAYG 240
EVNGLLNPAL VEVYGLSSTQ GSSTGAGGAG GNTGGDTNTQ TYARPGIGFK LPSTDSESSK 300
ATLITPGLAW TAQDVGNLVV SGTSLSFQLG GWLVTFTDFI KPRSGYLGLQ LTGLDANDSD 360
QRELIWAPPA LNRLSWQLGQ PLGPRGECVG FQGGVGGSSS VRLASSYKYH HRNEGYLIGA 420
40 HQCFGLSGEL YRPGFVQGFH SKLRPKPKHL PLPALGAGEK SRFLW 465
<212> Type : PRT
<211> Length : 465
SequenceName : SEQ ID 306
SequenceDescription :

45 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
50 MLGSIPVLVN RSGSDSNKFQ ATDQKWSYTD LQSDQTKLNL SAYGEVNGLL NPALVETYFG 60
TTRTSSTANQ NSTTVPGIGF KIPEQNNDK ATLITPGLAW TPQDVGNLVV SGTTVSFQLG 120
GWLVTFTDFV KPRAGYLGLQ LSGLNASDSD QRELIWAPRP WAAFRGSWVN RLGRVESVWD 180
LKGWADQAQ LAAQAATSST TTTATGATLP EHPNALAYQI SYTDKDSYKA STQSGGQTNQ 240
QNNSLYLHLI KPKKVESTTQ LDQGLKNLLD PNQVRTKLRQ SFGTDHSTQP QPQSLKTTTP 300
55 VFGAMSGNLG SVLSGGGAGG AGSTNSVDLS PVERVSGSLT INRNFSY 347
<212> Type : PRT
<211> Length : 347
SequenceName : SEQ ID 307
SequenceDescription :

60 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
65 MGQGGQSGTS AGNPDSLKQD KISKSGDSL TQDGNATGQQ EATNYTNLPP NLTPADWPN 60
ALSFTNKNNA HRAQLFLRGL LGSIPVLVNR SGSDSNKFQA TDQKWSYTDL QSDQTKLNL 120
AYGEVNGLLN PALVETYFGN TRAGGSGSNT TSSPGIGFKI PEQNNDKAT LITPGLAWTP 180

QDVGNLVVSG TSLSFQLGGW LVSFTDFIKP RAGYLGLQLS GLDASDSDQR ELIWAKRPWA 240
AFRGSWVNRL GRVESVWDLK GVWADQAQLA AQAATSEASG SALAPHPNAL AFQVSVVEAS 300
AYSSSTSSSG SGSSSNTSPY LHLLKPKKVE STTQLDQGLK NLLDPNQVRT KLRQSFQTDH 360
STQPQSLKTT TPVFGTSSGN IGSVLGGGA GGGSSSGSQS GVDLSPVERV SGH 413

5 <212> Type : PRT
<211> Length : 413
SequenceName : SEQ ID 308
SequenceDescription :

10 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

15 MGLQLSGLDA SDSDQRELIW AKRPWAAFRG SWVNRLGRVE SVWDLKGVWA DQAHSVSES 60
QAATSTTTT ATGDTLPEHP NALAYQISST DKDSYKASTQ GSGQNSQNT SPYLHLIKPK 120
KVTASDKLDD DLKNLLDPNE VRVKLRQSFQ TDHSTQPQPQ PLKTTTPVFG TNSGNLGSVL 180
SGGGTTQDSS TTNQLSPVQR VSGWLVGQLP STSDGNTSST NNLAPNTNTG NEVVGVGDLS 240
KRASIESSRL WIALKP 256

20 <212> Type : PRT
<211> Length : 256
SequenceName : SEQ ID 309
SequenceDescription :

25 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

30 MRDNTAKGIT AGSGSQTTY DPARTEATLT TTTFALRRYD LAGRALYDL FSKLNPQTPT 60
RDANCQITFN PFGGFGLSGS APQQWNEVKN KVPVEVAQDP TDPYRFAVLL VPRSVVYEQ 120
LQRGLALPNQ GSSSGSGQQN TTIGAYGLKV KNAEADTAKS NEKLQDESK SSNGSSSTST 180
TTQRGSTNSD TKVKALKIEV KKKSDSEDNG QLQLEKNDLA NAPIKRGEES QGSVQLKADD 240
FGTAPSSSGS GGNSNPGSPT PWRPWLATEQ IHKDLPKWSA SILILYDAPY ARNRTAIDRV 300
DHLDPKVM TA NYPPSWRMPK WNHHLWDWK ARDVLFTTQ FDESNTSNTK QGFQKEADSD 360
35 KSAPIALPFE AYFANIGNLT WFGQALLVFG GNGHVTKSAH TAPLSIWLYI YLVKAVTFRL 420
LLANSLLSKS NIYKKTAN 438
<212> Type : PRT
<211> Length : 438
SequenceName : SEQ ID 310
SequenceDescription :

40 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

45 MRDNIAGKIT AGSNTQTTY DPTRTEATLT TATTFALRRY DLAGRALYDL DFSKLNQPTP 60
TRDQTGQITF NPFGGFGLSG AAPQQWNEVK DKVPVEVAQD PSNPYRFAVL LVPRSVVYEQ 120
QLQRGLALPN QGSSSGSGQQ NTTIGAYGLK VKNAEADTAK SNEKLQGYES KSSNGSSSTS 180
TTQRGSSSNE NKVKALQVAV KKKSGSQGNS GDQGTQVEL ESNDLANAPI KRGSNNNQV 240
50 QLKADDFGTA PSSSGSGTQD GTPTPWPWL TTEQIHNDPA KFAASILILY DAPYARNRTA 300
IDRVHLDLPK VMTANYPPSW RTPKWNHGL WDWKARDVLL QTTGFFNPRR HPEWFDGGQT 360
VADNEKTGFD VDSENTKQG FQKEADSDKS APIALPFEAY FANIGNLTWF EQALLVFGIC 420
LS 422
<212> Type : PRT
<211> Length : 422
SequenceName : SEQ ID 311
SequenceDescription :

60 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

65 MLWPFWRVWW KRVLTSTQTRA PAKPNPLTVP PTCTWWSLRK LPNPTKLDDD LKNLLDPNEV 60
RARMLKSFGT ENFTQPQPQ QALKTTTPVF GTSSGNLGSV LSGGGYHAGL KHHQSTVTRS 120
TGEWVDR 127
<212> Type : PRT
<211> Length : 127

SequenceName : SEQ ID 312
SequenceDescription :

5 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :
10 MRDNSAKGIT AGSESQTTY DPTRTEAALT ASTTFALRRY DLAGRALYDL DFSRLNPQTP 60
 TRDQTGQITF NPFGGFGLSG AAPQQWNEVK NKVPVEVAQD PSNPYRFAVL LVPRSVVYYE 120
 QLQRGLALPN QGSSSGSGQQ NTTIGAYGLK VKNAEADTAK SNEKLQDES KSSNGSSSTS 180
 TTTQRGGSSG DTKVKALQVA VKKSGSQGN SGEQTEQVE LESNDLANAP IKRGEESGQS 240
 VOLKAADFGT TPSSSGSGGN SNPGSPTPWR PWLATEQIHK DLPKWSASIL ILYDAPYARN 300
 RTAIDRVDHL DPKVMTANYP PSWRTPKWNH HGLWDWKARD VLLQTTGFFN SRRHPEWFDQ 360
 GQAVADNTQT GFDTDDTDNK KTRLKSGSWL RQAGPDEPPV WSVLRQHWQP HLVRSASAFGV 420
15 WDLFVLIN 428

 <212> Type : PRT
 <211> Length : 428
 SequenceName : SEQ ID 313
 SequenceDescription :

20 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :
25 MFGLKVKNAE ADTAKSNEKL QGAEATGSST TSGSGQSTQR GGSSGDTKVK ALQVAVKKKS 60
 GSQGNSGDQG TEQVELESND LANAPIKRGs NPASPTQGSR LRHHPIQFGI WSIRHPHPLK 120
 AVACDRANSQ GPPQMIRLDP HSVRCALCL 149

 <212> Type : PRT
 <211> Length : 149
30 SequenceName : SEQ ID 314
 SequenceDescription :

 Sequence

35 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :
 MFGLKVKDAT VDSSKQSTES LKGEESSSSS TTSSTSTTQR GGSSGDTKVK ALQVAVKKKS 60
 DSEDNGQIEL ETNNLANAPI KRGSNNNQQV QLKADDFGTS PSSSESGQSG TPTPWTPWLA 120
 TEQIHKDLPK WSASILILYD APYARNRTAI DRVDHLDPKV MTANYPPSWR TPKWNHHGLW 180
40 DWKARDVLVQ TTGFFNPRRH PDWFDQQAQV AENTQTGFDT DDTDNKKQGF RKQGEQSPAP 240
 IALPFEAYFA NIGNLTWFGQ ALLVFGICLS 270

 <212> Type : PRT
 <211> Length : 270
 SequenceName : SEQ ID 315
45 SequenceDescription :

 Sequence

50 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :
 MGSQNQGSTT TTSAGNPDSL VTDKVDQKGQ VQTSGQNLSD TNYTNLSPNF TPTSDWPNAL 60
 SFTNKNNQR AQLFLHGLLG SIPVLVNKSG ENNEKFQATD QKWSYTELKS DQTKLNLPAY 120
 GEVNGLLNP LVTYFGTTR TSSTANQNST TVPGIGFKIP EQNNSKAVL ITPGLAWTPQ 180
 DVGNLVVSQT SFSFQLGGWL VSFTDFVKPR AGYLGLQLTG LDASDATQRA LIWAPPALSG 240
55 LSWQLGQPVG PRGECVGFEG GVGSSSSVRL ARIYHHRNRG YLTGAPECFG LSGECGGSEC 300
 LQAKHELPRN PIH 313

 <212> Type : PRT
 <211> Length : 313
 SequenceName : SEQ ID 316
60 SequenceDescription :

 Sequence

65 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :
 MSFGLVGTVN NNGWKSPFRH ETKYRAGYDK FKYYKTHYRG AKKAGTNDNR WRWTAWFDLD 60
 FAHQKIVLIE RGELHRQADL KKSDPATNET SKTVWGSIKE KLLQNVNNLH SEKGVFLWFR 120

QSGFTTTRN
<212> Type : PRT
<211> Length : 129
SequenceName : SEQ ID 317
SequenceDescription :
5
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
10 <400> PreSequenceString :
MAEPLAVDPT GLSAAAAKLA GLVFPQPPAP IAVSGTDSVV AAINETMPST ESLVSDGLPG 60
VKAALTRTAS NMNAAADVYA KTDQSLGTSI SQYAFGSSGE GLAGVASVGG QPSQATQLLS 120
TPVSQVTTQL GETAAELAPR VVATVPQLVQ LAPHAVQMSQ NASPIAQTIS QTAQQAAQSA 180
QGGSGPMPAQ LASAEKPATE QAEPVHEVTN DDQGDQGDVQ PAEVVAAARD FGAGGAGGQQ 240
15 PGGGVPAQAM DTGAGARPAA SPLAAPVDPS TPAPSTTTTL 280
<212> Type : PRT
<211> Length : 280
SequenceName : SEQ ID 318
SequenceDescription :
20
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
25 MRYLIATAVL VAVVLVGWPA AGAPPSCAGL GGTVQAGQIC HVHASGPKYM LDMTFPVDYP 60
DQQALTDYIT QNRDGFVNVA QGSPLRDQPY QMDATSEQHS SGQPPQATRS VVLKFFQDLG 120
GAHPSTWYKA FNYNLATSQP ITFDTLFVPG TPLDSIYPI VQRELARQTG FGAAILPSTG 180
LDPAHYQNFA ITDDSLIFYF AQGELLPSFV GACQAQVPRS AIPPLAI 227
<212> Type : PRT
30 <211> Length : 227
SequenceName : SEQ ID 319
SequenceDescription :

Sequence
35 -----
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKMVKSIAAG LTAAAAIGAA AAGVTSIMAG GPVVYQMOPV VFGAPLPLDP ASAPDVPTAA 60
QLTSLNLSLA DPNVSFANKG SLVEGGIGGT EARIADHKLK KAAEHGDLPL SFSVTNIQPA 120
40 AAGSATADVS VSGPKLSSPV TQNVTFVNQG GWMLSRASAM ELLQAAGN 168
<212> Type : PRT
<211> Length : 168
SequenceName : SEQ ID 320
SequenceDescription :
45
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
50 MTYSPGNPGY PQAQAGSYG GVTPSFAHAD EGASKLPMYL NIAVAVLGLA AYFASFGPMF 60
TLSTELGGGD GAVSGDTGLP VGVALAALL AGVALVPKAK SHVTTVAVLG VLGVFLMVSA 120
TFNKPSAYST GWALWVVLAF IVFQAVAAVL ALLVETGAI APAPRPKFDY YGQYGRYQY 180
GQYGVQPGGY YGQQAQQA GLQSPGPQQS PQPPGYGSQY GGYSSSPSQS GSGYTAQPPA 240
QPPAQSGSQ SHQGPSTPPT GFPSFSPPPP VSAGTGSQAG SAPVNYSNPS GGEQSSSPGG 300
55 APV 303
<212> Type : PRT
<211> Length : 303
SequenceName : SEQ ID 321
SequenceDescription :
60
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
65 MKCPGVSDCV ATVRHDNVFA IAAGLRWSAA VPPLHKGDAV TKLLVGAIAG GMLACAAILG 60
DGIASADTAL IVPGTAPSPY GPLRSLYHFN PAMQPQIGAN YYNPTATRHV VSYPGSFWPV 120
TGLNSPTVGS SVSAGTNNLD AAIRSTDGPI FVAGLSQGTI VLDREQARLA NDPTAPPPGQ 180

LTFIKAGDPN NLLWRAFRPG THVPIIDYTV PAPAESQYDT INIVGQYDIF SDPPNRPGNL 240
LADLNAIAAG GYYGHSATAF SDPARVAPRD TTTTNSLGA TTTYFIRTD QLPLVRALVD 300
MAGLPPQAAG TVDAALRPII DRAYQPGPAP AVNPRDLVQG IRGIPAIAPA IAIPIGSTTG 360
ASAATSTAAA TAAATNALRG ANVGPGANKA LSMVRGLLPK GKKH 404

5 <212> Type : PRT
<211> Length : 404
SequenceName : SEQ ID 322
SequenceDescription :

10 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSTLSTT... LPPPFDAIPN PIEDLDVLVA LQVWFAAGSI GVSAAQLCEI TPEHDA...Q 60
15 KAPHCPAESD QTPAGAAGDG DLPEVGGRVT SPPQPPVAAL TGYSANIGGL SVPHSWNLPP 120
AVRQVAAMFP GATPMYMTGS SDGSYAGLAA AGLAGTGLAG LAARGGSAPT PAAAAPAGAG 180
GAGPAATRPA AQQTPAVPAA AAGSAIPGLP PGLPPGVVAN LAATLAAIPG ATTIIVPPSP 240
NANQ 244
<212> Type : PRT
20 <211> Length : 244
SequenceName : SEQ ID 323
SequenceDescription :

25 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MDVALGVAVT DRVARLALVD SAAPGTVIDQ FVLDVAEHPV EVLTETVVG T DRSLAGENHR 60
LVATRLCWPD QAKADELQHA LQDSGVHDVA VISEAQATA LVGA AHAGSA VLLVGDETAT 120
30 LSVVGD PDAP PTMVAVAPVA GADATSTVDT LMARLGDQAL APGDVFLVGR SAEHTTVLAD 180
QLRAASTMRV QTPDDPTFAL ARGAMAAGA ATMAHPALVA DATTSLPRAE AGQSGSEGEQ 240
LAYSQASDYE LLPVDEYEEH DEYGAAADRS APLSRRSLLI GNAVVAFAVI GFASLAVAVA 300
VTIRPTAAEK PVEGHQNAQP GKFMPLLP TQ QQAPVPPPP DDPTAGFQGG TIPAVQNVVP 360
RPGTSPGVGG TPASPAPAP AVPGVVPAPV PIPVPIIIPP FPGWQPGMPT IPTAPP TTPV 420
35 TTSATTPPTT PPTTPVTTP TTPPTTPVT PPTTPPTTPV TTPPTTVAPT TVAPTTVAPT 480
TVAPT TVAPA TATPTTVAPQ PTQQFTQQPT QQMFTQQQTV APQTVAFAPQ PPSGGRNGSG 540
GGDLFGGF 548
<212> Type : PRT
<211> Length : 548
40 SequenceName : SEQ ID 324
SequenceDescription :

45 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKNARTTLIA AAIAGTLVTT SPAGIANADD AGLDPNAAAG PDAVGFDPNL PPAPDAAPVD 60
TPPAPEDAGF DPNLPPPLAP DFLSPPAEEA PPVPVAYSVN WDAIAQCESG GNWSINTGNG 120
YYGGLRFTAG TWRANGSGSGS AANASREEQI RVAENVLRSQ GIRAWPVCGR RG 172

50 <212> Type : PRT
<211> Length : 172
SequenceName : SEQ ID 325
SequenceDescription :

55 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
60 MTRLIPGCTL VGLMLTLLPA PTSAAGSNTA TTLFPVDEVT QLEHTFLDC HPNGSCDFVA 60
GANLRTPDGP TGFPPGLWAR QTTEIRSTNR LAYLDAHATS QFERVMKAGG SDVITTVYFG 120
EGPPDKYQTT GVIDSTNWST GQPMTDVNI VCTHMQVVYP GVNLTSPSTC AQANFS 176

<212> Type : PRT
65 <211> Length : 176
SequenceName : SEQ ID 326
SequenceDescription :

Sequence

5 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTPGLLTITAG AGRPRDRCAR IVCTVFIETA VVATMFVALL GLSTISSKAD DIDWDAIAQC 60
ESGGNWAANT GNGLYGGLQI SQATWDSNGG VGSPAAASPQ QQIEVADNIM KTQGP GAWPK 120
CSSCSQGDAP LGS LTHILTF LAAETGGCSG SRDD 154
<212> Type : PRT
10 <211> Length : 154
SequenceName : SEQ ID 327
SequenceDescription :

Sequence

15 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MMQQA VSGIT GALGGAVGGV MGPLTQLPQQ AMQAGQGAMQ PLMSALQQTY GAEGLDVADG 60
ARLVDSIEGE PGLGGEPGAG DVGAGGGGGG TTPTGYLGPP PVPTSSPPTT PAGAPAKSVT 120
20 PDPVSGTPRA SGPAGMTGMP MVPPGALGAG AEGANKDKPV EKRVTGCAEW STGQGPLNST 180
AECSGEICRR QAGGHQVDAT DPCCAERRQG 210
<212> Type : PRT
<211> Length : 210
SequenceName : SEQ ID 328
25 SequenceDescription :

Sequence

30 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MIRELVT TAA ITGAAIGGAP VAGADPQRYD GDVPGMNYDA SLGAPCSSWE RFIFGRGPSG 60
QAEACHFPPP NQFP PAETGY WVISYPLYGV QQVGAPCPKP QAAAQSPDGL PMLCLGARGW 120
QPGWFTGAGF FPPEP 135
<212> Type : PRT
35 <211> Length : 135
SequenceName : SEQ ID 329
SequenceDescription :

Sequence

40 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKTGTGTTIKL GIVWLVLVSVF TVMIIVVFGQ VRFHHTTGYS AVFTHVSGLR AGQFVRAAGV 60
EVGKVAKVT L IDGDKQVLVD FTVDRSLSLD QATTASIRYL NLIGDRYLEL GRGHSGQRLA 120
45 PGATIPLEHT HPALDL DALL GGFRPLFQTL DPKVNSIAS SIITVFQGGG ATINDILDQT 180
ASLTATLADR DHAIGE VVNN LNTVLATTVK HQTEFDRTVD KLEVLITGLK NRADPLAAAA 240
AHISSAAGTL ADLLGRIVHC CTAASGTSRA SSSRS 275
<212> Type : PRT
<211> Length : 275
50 SequenceName : SEQ ID 330
SequenceDescription :

Sequence

55 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTPRSLVRIV GVVVATTLAL VSAPAGGRAA HADPCSDIAV VFARGTHQAS GLGDVGEAFV 60
DSLTSQVGGR SIGVYAVNYP ASDDYRASAS NGSDDASAHI QRTVASCPNT RIVLGGYSQG 120
ATVIDLSTSA MPPAVADHVA AVALFGEPSG GFSSMLWGGG SLPTIGPLYS SKTINLCAPD 180
60 DPICTGGGNI MAHSVYVQSG MTSQAATFAA NRLDHAG 217
<212> Type : PRT
<211> Length : 217
SequenceName : SEQ ID 331
SequenceDescription :
65

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MISTTRIDFL WILSVAFASM IALATLLTLI NQVVGTPYIP GGDSPAGTDC SELASWVSNA 60
ATARPVFGDR FNTGNEEAAL AARGFQQGTA PNALVIGWNG HHTAVTLPDG TPVSSGEGGG 120
5 VRVGGGGAYQ PKFTHHMYLP MDVDAGEDQP PAPDEPVTAV DDVEPEMPAP CPTQRPPVTP 180
RHNLCKNLRT MPGALSAALA AAPVWPAPI SGRGFSTSL LAKRNHPVIV GK 232

<212> Type : PRT
<211> Length : 232
10 SequenceName : SEQ ID 332
SequenceDescription :

Sequence

15 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTTMITLRRR FAVAVAGVAT AAATTVT LAP APANAADVYG AIAYSGNGSW GRSWDYPTRA 60
AAEATAVKSC GYSDCKVLTS FTACGAVAN DRAYQGGVGP TLAAAMKDAL TKLGGGYIDT 120
WACN 124

20 <212> Type : PRT
<211> Length : 124
SequenceName : SEQ ID 333
SequenceDescription :

25 Sequence

30 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MAGLNIYVRR WRTALHATVS ALIVAILGLA ITPVASAATA RATLSVTSTW QTGFIARFTI 60
TNSSTAPLTD WKLEFDLPAG ESVLHTWNST VARSGTHYVL SPANWNRIIA PGGSATGGLR 120
GGLTGSYSPP SSCLLNGQYP CT 142

<212> Type : PRT
<211> Length : 142
SequenceName : SEQ ID 334
35 SequenceDescription :

Sequence

40 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MLTRAIKTQL VLLTVLAVIA VVVLGWYFLR IPSLVGIGRY TLYAELPRSG GLYRTANVTY 60
RGITIGKVTG VEPTERGARA TMSIDNGYQI PTDASANVHS VSAVGEQFVD LVSTRTS GPY 120
LRHGQTITTT TVPSQIGPAL DAANRGLAVL PKDRVASVLH EASEAVGGLG SSLNRLIEAT 180
QAIAHDVRGS LEDIDDIER SAPIIDSQVN SGNEIARWAA NLNTLAAQTA QTDPAVRSIL 240
45 ANAAPTADQV NATFSDVRES LPQTLANLEV VIDMLKRYHN GVEQALVFLP QSGAIAQSVT 300
TEFPGQAGLG VGGLALNQPP PCLTGFLPAS EWRSPADTST APLPKGTYCR IPMDASNVVR 360
GARNNPCVDV PGKRAATPRE CRSNEAYVPG GTNPWYGDPN QMLSCPAPAA RCDQPVKPGQ 420
VIPAPSVNNG INPLPADQLP GTPPPVNDPL QRP GSGTVQC NGQQPNPCVY TPSTFPTTIY 480
DVQSGKVVAP DGVVYSVEAS THAGADGWKV MLAPTG 516

50 <212> Type : PRT
<211> Length : 516
SequenceName : SEQ ID 335
SequenceDescription :

55 Sequence

60 <213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MLNNTQFLNL MKSYMKPEFY MSSIKNTTNL DLSSITNTIQ KAMNIFFTTN KISTESMQSL 60
FKKNSEIIQN NINTILNSTK EVINSKDFKQ ATEYHQKCVK SIYETSMDNA KELANIAYEA 120
SNKIFEAAANK HITKNIHNAS NNIHNTAEQV QKNFNKSA 159

<212> Type : PRT
<211> Length : 159
SequenceName : SEQ ID 336
65 SequenceDescription :

Sequence


```
-----
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
5  MNIKLVITYFL ILVSSLKVNA DLNHIQDSFK YQEAEQLTIE LPWNDCTAIH KFLLEKLFFS      60
   EQQIKKENKI HEKYKQFYLO HNNKLSDFSM QFLEKKSEIN SVETLISGFL KFCEDNFQTS      120
   KSKSHSLNFF QKQODQWLHN IRNENYKTTY KKKYEDNTFR NIN                        163
<212> Type : PRT
<211> Length : 163
      SequenceName : SEQ ID 337
10      SequenceDescription :

Sequence
-----
<213> OrganismName : Rickettsia prowazekii
15 <400> PreSequenceString :
   MKKLLLIATA SATILSSSVS FAECIDNEWY LRADAGVAMF NKEQDKATGV KLKSNKAIPi      60
   DLGIGYYISE NVRADLTIGT TIGGKLKKYG AATNTHFTGT NVSVSHKPTV TRLLINGYVD      120
   LTSFDMFDVF VGGGVGPALV KEKISGVSGI ASNTKNKTNV SYKLIFGTSA QIADGVKVEL      180
   AYSWINDGKT KTHNVMYKGA SVQTGGMRYQ SHNLTGVGRF GI                        222
20 <212> Type : PRT
<211> Length : 222
      SequenceName : SEQ ID 338
      SequenceDescription :

Sequence
-----
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
30 MKKNMRKQML KIISIIIIISL LLSSCSESTR DENGLLTDSQ STIIRDYIIS QNSKNLKVNL      60
   KEKFGSNLKG VKLIGIKLTN EDLSGIDFTS CEILRTDFMG SNLEKAILTN SVIQESNFAD      120
   SVIKNISGYN ADFQGSIFNN ITLQNTNFVQ SNFSDTAFNK STIINVNFEN SKFSNVLWCH      180
   SNIDSSNFQK THLKNNSEFKN TNVMNSIFYG ADLGKSVINN TNFTNNYFES SDLSNTKFTS      240
   VLIKDSNFTQ SIFNSVNFNN IQSNNSFFSY TSFEDSTLHN IHLTKCDLQN STINSSVFNN      300
   FKIDNAILTN MSLNDNTFNN LSIKNSNTNF VRINKSKGFN ITLLNTNYSN AIFSNNDLKE      360
35 FKVINTDLNN SEIINSNFTN GQFNNVNFSQ SLIQNVNFTD VKITLGNLNO VALINSNLIN      420
   TNIINSVLSN SQINNINYQA YYSFINTNVS NNIVINDNSN QIPPNIVIN SEKDLQNISN      480
   LANMNLTNFN LSNLVFNGVD FSKSIFKKAN LTNTVIKNSI LKDNANFSAI LTKTDFSKSI      540
   LTGSIFKFAQ IDQTCFSNSD LTNTDFTEAT IKNTAFDNAN THGIKGLE                    588
40 <212> Type : PRT
<211> Length : 588
      SequenceName : SEQ ID 339
      SequenceDescription :

Sequence
-----
45 <213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
   MIQKFTNVKL NDMRKILSFL MMCSLHLGLQ SQTWHGDPDS VAALPSIGIQ ESSCTRITFE      60
   VVFPGFYSVE KREGNQVFQR ISMPGCGSFG NLGEAELPVL KKMIAVPEFS TANVAVKIKE      120
50 TETFDNYNIY PNPTYVVEEL PEGGTYLVEA FAINNDYYSQ NVSLPSTHYV YSQDGYFRSQ      180
   RFIEVTLYPF RYNPVRQEIL FAKKIEVTIT FDNPPPLQK NTGIFNKVAS SAFINYEADG      240
   KSAIENDMVF SRGTTYISG NVASNLPONC DYLVYDDMF NVNQPHDEI KRLCEHRAFY      300
   NGFDVAAVSI KDVLNSFSPN ATSYINETKL KNFIRSVYNQ SNAKRTLDGK LGYVLLIGKP      360
   LSKYLADTDN TKVPTSFHFN VSLIPSHPTF GSICASYFF SCVSPLDTVG DLFGRFSVT      420
55 NAHELHNLIE KTINKEISYN PIAHKNILYA EGKGCDAPIL RLFLKEIASG YTVNSILKSN      480
   QVSAIDSIFD CLNNGSHHFY FNTHGMPTVW GIGQGLDVNT LTARLNNTSS QGLCTSLSCS      540
   SAVADSTIRS LGEVLTTYAP NKGFSAPLGG SRATQYAVYL EGPCPPSEFY EYLPYSLYHN      600
   LSTVVGEMLL SSIINTNSVD TYSKFNFNLL GDPALNIMAH GMEVSNCTIL PNNTIISPI      660
   TIKNGGCLKI PEKGVLFHTN NGSIQVMSGG TLEIGNQAKI SGETGANPTF ITVYGDGLAI      720
60 NKQVEIDNID RLNLFTSTHSV MPKFHFDSVK FNSAPLYTTN CIVEISNCEF TNRSDIISKI      780
   CDLSVENSMT SSSGITVFKP MATSSITGLS TKAKITDNTF FATGNFAYHI TNTPGLTATS      840
   NAAIKLDNIP EYYISGNKIV NCDEALVLNN SGNRTNRLHN ITRNVIKNCR IGSTLYNSYG      900
   IYNRNKISNN HIGVRLNNS CFYFDNAPVI NEEDKQTFIS NRTWQLYSSN GTFPLNFHYN      960
   SLQGGDTDTW IYNDTYTNRV IDVSNNHWGN NDLEDPNQVF NTPDLFIWIP FWDGLPNGRS      1020
65 GNSSAEAVEF QTALDCIGNS DYLSAKVALK MMVETYPESD FAIAALKELF RIEKMSGNDY      1080
   EGLKDYFRSN PTIISQNLFF PTADFLSARC DIVCENYQSA IDWYENRLNS EISYQDSVFA      1140
   VIDLGDIYWN MQLDSLRTGT IDLNILSCEQ RKSLESHQNV KNYLLSTLPE STGTLLPPE      1200
```

CNKSSLDKSK IISISPNPAK AVVTIIYYTD NPSCSVIKIY GINGASADIT GLPKHLSEGY 1260
YSIQFNTSNF DPGFYLVTLN VDQKIIDTEK LRIK 1294
<212> Type : PRT
<211> Length : 1294
5 SequenceName : SEQ ID 340
 SequenceDescription :

Sequence

10 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
 MQGKNTIVTT GDYSIGLLSQ TSGNLNTDTI IRVNSDGSVT PSFSDGDDTF IVTAGNHAVG 60
 VLACASPGSA CACVSSLDEE STADTGSNEN NAIAKLDMAK GEITTHGTES YAAYANGTVV 120
 KAGDTLDYTN ASVTLTDVDI TTHGDNAHAI AARQGTVSFN QGEIYTTGPD AAIAKIYNGG 180
15 TVTLKNTSAV AHQSGGIVLE SSINGQEATV DILSGSSLRS ANEILYHKDE TSNVTITDSE 240
 VSSAADVFIN NIKGHLTVDA TNSKITGSAN ISTDDNTHTY LSLSDNSTWD IKADSTVSNL 300
 TVDNSTVYIS RADGRDVEPT RLITITENYVG NNGVLHLRTE LDDDNSATDK VVINGNTSGT 360
 TRVKVTNAGG SGAYTLNGIE IISVEGESNG EFIKDSRIFA GAYEYSLTRG NTEATNKNWY 420
 LTNFQATSGG ETNSGGSSAP TVAPTFLVRP EAGSYVANLA AANTLFVMRL NDRAGETRYI 480
20 DPVTEQERSS RLWLRQIGGH NAWRDSNGQL RTTSHRYVSQ LGGDLTLTGGF TDSDSWRLGV 540
 MAGYARDYNL THSSVSDYRS KGSVRGYSAG LYATWFADDI SKKGAYIDSW AQYSWFKNSV 600
 KGDELAYESY SAKGATVSLE AGYGFALNKS FGLEAAKYTW IFQPQAQAIW MGVDHNAHTE 660
 ANGSRIENDA NNNIQTRLGF RTFIRTQEK N SGPHGDDFEP FVEMNWIHNS KDFAVSMNGV 720
 KVEQDGVSNL GEIKLGVNGN LNPAASVWGN VGVQLGDNGY NDTAVMVGLK YKF 773
25 <212> Type : PRT
 <211> Length : 773
 SequenceName : SEQ ID 341
 SequenceDescription :
30 Sequence

 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
35 MTKLKLLALG VLIATSAGVA HAEGKFSLGA GVGVEHPYK DYDTDVYPVP VINYEGDNFW 60
 FRGLGGGYL WNDATDKLSI TAYWSPLYFK AKDSGDHQMRL HLDDRKSTMM AGLSYAHFTQ 120
 YGYLRITTLA DTLDNSNGIV WDMAWLYRYT NGGLTFTPGI GVQWNSNQ N EYYYGVSRKE 180
 SARSGLRGYN PNDSWSPYLE LSASYNFLGD WSVYGTARYT RLSDEVTDSP MVDKSWTGLI 240
 STGITYKF 248
40 <212> Type : PRT
 <211> Length : 248
 SequenceName : SEQ ID 342
 SequenceDescription :
45 Sequence

 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
50 MKKIALAGLA GMLLVASAVN AMSISQAGK EYTNIGVGFG TESTGLALSG NWTNDDDDGD 60
 VAGVGLGLNL PLGPLMATVG GKGVYTNPNY GDEGYAAAVG GGLQWKIGNS FRLFGEYYYS 120
 PDSLSSGIQS YEEANAGARY TIMRPVSIEA GYRYLNLSGK DGNRDNAVAD GLYVGVNASF 180
 <212> Type : PRT
 <211> Length : 180
55 SequenceName : SEQ ID 343
 SequenceDescription :

Sequence

60 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
 MTTLTARVFT TAEIIRKTV IALVCHLNCS RQETVTMNTK IMALAIMMAS FAANASVLPE 60
 TPVPFKSGTG AIDNDTVYIG LGSAGTAWYK LDTQAKDKKW TALAAPGGP REQATSAFID 120
 GNLYVFGGIG KNSGLTQVF NDVHKYNPKT NSWVKLMSHA PMGMAGHVTF VHNGKAYVTG 180
65 GVNQNIFFNGY FEDLNEAGKD STAIKINAH YFDKKAEDYF FNKFLLSFDP STQQWSYAGE 240
 SPWYGTAGAA VVNKGDKTWL INGEAKPGLR TDAVFELDFT GNNLKWNKLD PVSSPDGVAG 300
 GFAGISNDSL IFAGGAGFKG SRENYQNGKN YAHEGLKKS Y STDIDLWHNG KWDKSGELSQ 360

GRAYGVSLPW NNSLLIIGGE TAGGKAVTDS VLISVKDNKV TVQN 404
<212> Type : PRT
<211> Length : 404
SequenceName : SEQ ID 344
5 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
10 <400> PreSequenceString :
MATGGAALAG KAVMGAAAGA AGGASALQAA FQKASASMET GGDMSMGSV VSSGGNGGGE 60
AGTAGSSPFA QAAGFGDSGS SSSGGGFAKA AKLATGTASE LAKGVGSQVK QGFQERVSET 120
TGGKLAASIR ESMPEKEASQ SGQFEGNSLG ADSGPDSNEV RS 162
<212> Type : PRT
15 <211> Length : 162
SequenceName : SEQ ID 345
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRVLIPGVI LCGADVAQAV DDKNMYMYFF EEMTVYAPVP VPVNGNTHYT SESIERLPTG 60
NGNISDLLRT NPAVRMDSTQ STSLNQGDIR PEKISIHGAS PYQNAYLIDG ISATNNLNPA 120
25 NESDASSATN ISGMSQGYL DVSLLDNVTL YDSFVPVEFG RFNGGVIDAK IKRFNADDSK 180
VKLGYRTRL DWTSHIDEN NKSAFNQSS GSTYFSPDFK KNFYTLNFNQ ELADNFGVTA 240
GLSRRQSDIT RADYVSNDGI VAGRAQYKNV IDTALSKFTW FASDRFTHDL TLKYTGSSRD 300
YNTSTFPQSD REMGNKSYGL AWDMDTQLAW AKLRTTVGWD HISDYTRHDH DIWYTELSCT 360
YGDITGRCTR GGLGHISQAV DNYTFKTRLD WQKFAVG DVS HQPYFGAEYI YSDAWTERHN 420
30 QSESYVINAA GKKTNHTIYH KGKGS LGIDN YTYLMADHIS WRNVSLMPGV RYDYDNYLSN 480
HNISPRFMTE WDIFADQ TSM ITAGYNRYYG GNILDMGLRD IRNSWTESVS GNKTLTRYQN 540
LKTPYNDELA MGLQQKIDKN VIARASEAHD QISKSSRTDS ATKTTITEYN NDGKTKTHSF 600
NLSFELAEPL HIRQVDINPQ IVFSYIKSKG NLSLNNGYEE SNTGDNQVVY NGNLVSYSV 660
PVADFNPLK ISLNMDFTHQ PSLVWANTL AWQEAR KARI ILGKTNAQYI SEYSDYKQYV 720
35 DEKLDSSLTW DTRLSTWTPQF LKQONLTISA DILNVLD SKT AVDTTNTGVA TYASGRTFWL 780
DVSMKF 786
<212> Type : PRT
<211> Length : 786
SequenceName : SEQ ID 346
40 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
45 <400> PreSequenceString :
MKKTLLA IML AGTAFASQAG TLVSQGTEAS ANLTLTKPIV VNNTIQPVKG VYSGTLTAWT 60
PLATGIVGAS DGQSHDYAVT FPDDIYAESS TSADAVISGD NNPDKHLKVS LITTLEQDPPS 120
AASEEIGGKR YMLKNTGTG GAYRVVSHMK EQVVEPDSYT IRTQAYIYAE 170
<212> Type : PRT
50 <211> Length : 170
SequenceName : SEQ ID 347
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MGIYHWSRKT KMKRTKSIRH ASFRKNWSAR HLT PVALAVA TVFMLAGCEK SDETVSLYQN 60
ADDCSAANPG KSAECTTAYN NALKEAERTA PKYATREDCV AEFGEQGCQQ APAQAGMAPE 120
60 NQAQAQSSG SFWMLMAGY MMGRMLMGGA GFAQQPLFSS KNPASPAYGK YTDATGKNYG 180
AAQPGRMTV PKTAMAPKPA TTTTVTRGGF GESVAKQSTM QRSATGTSSR SMGG 234
<212> Type : PRT
<211> Length : 234
65 SequenceName : SEQ ID 348
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
5 MTKMSRYALI TALAMFLAGC VGQREPAPVE EVKPAPEQPA EPQQPVPTVP SVPTIPQQPG 60
PIEHEDRTAP PAPHIRHYDW NGAMQPMVSK MLGADGVTAG SVLLVDSVNN RTNGSLNAAE 120
ATETLRNALA NNGKFTLVSA QQLSMAKQQL GLSPQDSLGT RSKAIGIARN VGAHYVLYSC 180
ASGNVNAPTL QMQLMLVQTG EIIWSGKGAV SQQ 213
<212> Type : PRT
10 <211> Length : 213
SequenceName : SEQ ID 349
SequenceDescription :

Sequence

15 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MTKLMQFVQR CYYMTNKKMY FILILVFTLL QVCFFALWKA RDGSTTSLEC TSTLTRNAKT 60
DHSLYYSANL SVILKKDGSF SFTIVGLTDE DTPRKFSHSY FFTYKIDSNG RISGNAKAKV 120
20 SGLLENQIKDE NFRLNFLDAS LTGKGNARLS KFNNVYIFSI PGLIINTCAP I 171
<212> Type : PRT
<211> Length : 171
SequenceName : SEQ ID 350
25 SequenceDescription :

Sequence

30 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MGRISSGGMM FKAITTVAAL VIATSAMAQD DLTISSLAKG ETTKAAFNQV VQGHKLPWV 60
MKGGTYTPAQ TVTLGDETYQ VMSACKPHDC GSQRIAMWS EKSNOQTGLF SAIDEKTSQE 120
KLTWLVNDA LSIDGKTVLF AALTGSLENH PDGFNFK 157
<212> Type : PRT
35 <211> Length : 157
SequenceName : SEQ ID 351
SequenceDescription :

Sequence

40 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKQFLEKAV FTVAATAATV VLGKMKADAD TYTLQEGDSF FSVAQRYHMD AYELASMNGK 60
DITSLILPGQ TLTVNGSAAP DNQAAAPTD TQATTETNDA NANTYPVGQC TWGVKAVATW 120
45 AGDWWGNGGD WASSASAQGY TVGNTPAVGS IMCWTDDGGY HVAYVTAVGE DGKVQVLESN 180
YKDQQWVDNY RGWFDPNNSG TPGSVSYIYP N 211
<212> Type : PRT
<211> Length : 211
SequenceName : SEQ ID 352
50 SequenceDescription :

Sequence

55 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MSIKNILENK TTTIKVSFAG IATAASLILP MAVQAETTYT VKSGDTLSEI ASTHGTTVDK 60
LAKLNKINNI HLIHAGQILE LDAATEDTDA TPVQESQINE AETSASAKTS QTSEVTTTAP 120
VQESQTSEVI TSAPAETSQT SEVPTEANQT NEVSSAVSVE TSQTSEATTS APVETSQTSE 180
ATTAEPTEK TSQTNEVAAS AEENQTTST SGLSTSDAAA KEFIAQKESG GNYNAKNGQY 240
60 YGRYQLSDSY LNGDLSEENQ ERVADAYVSS RYGSWTAAQA FWNANGWY 288
<212> Type : PRT
<211> Length : 288
SequenceName : SEQ ID 353
SequenceDescription :
65

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKCQAFEDFK ATSLNKLSYT TGGATDGEII ANRMLQ GKAT KGEITMYTWN IIQNGWVNSL 60
VSWGIGGYNS SIGYSAQGNR GFSNYPYDVS MDSDNSSSSS NTTGGYVNYN QSFNSGW 117

5
<212> Type : PRT
<211> Length : 117
SequenceName : SEQ ID 354
SequenceDescription :

10
Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
15 MRYSQLCRKS LALLATGMIL TTSTLPSISI LAEDSTGAPA RPDGQAPAGG GANTTTYDYS 60
GINSGLVAN GSKVTSSSKT KSTTSAQNTA LVQNGGSLTL HKANLIKSGD DNNGDNDNFY 120
GINSILLAVN ERSKAYVSNS KLKASSSGSN GIFATDKATI YANKTSIATT ADNSRGLDAT 180
YNGNIIANKM AISTKGAHSA AIATDRGGGN ISTTNSSLNT SGSGSPLLYS TGNIQVNHVT 240
GTSSNSQIAG MEGLNTILIH NSNLISTMTN KTASDPIANG VIIYQSQSGD AEATTGQSAH 300
20 FELSKSKLTS SITSGSMFYL TNSANIILN QSTLNFDANK AKLLTVAGNS ANNWGTPGSN 360
GATVNFTGHK QTLKGDVDVD SISTLNMYLL DKTNYTGKTA VSTNSTNISP STSPITMNIS 420
KNSKWLTGH STVTNLNAEK GAKIVDKDGK TVSVISSSGQ KLVKGKSKYS LTVTGYSQK 480
VTTSSSNKPS SSYINRSDFD NYFKTTTAFV NNTKNTSN 518
<212> Type : PRT
25 <211> Length : 518
SequenceName : SEQ ID 355
SequenceDescription :

30
Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MNKIGDILRD ARIEKKLSFD DVVDKTGIAP HYILAMELDQ LKLLPEGKTN EYLEKYAHAV 60
GLDPVSIHIG YRNQEMSDEL ILPSSAELAA SSDSNIEKKN EGKSIEEPQE LAIDSLDVTQ 120
35 NITEETPQIE DFKVESEEAS KKIEKIPSRK SKYDYDEEPK KKFPWALILL ILLALTIISY 180
VGYVVYNQLQ TDSNKTELST STKSKSDTKN DANSTTQSQT SITTDFA DGG NNITLSNTNG 240
KVEVTFTLTG DEESWVSATN TTDGESGTTL TATDKTYTVT LAEGSTTSM L TVGSPSGVEI 300
TINGQKVDTT NLVNAGLTNI NLTVQ 325
<212> Type : PRT
40 <211> Length : 325
SequenceName : SEQ ID 356
SequenceDescription :

45
Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKSRKRQRKG LVRKNEIIL TLFVASAVSL LAFTNSFGVL AKSLHLEKIN KSITISLPFG 60
KKKMEQTARY YSGEQVQISS SAKKDSLKGK LSHYQNWIGT VKKIKSQKDS RQKHHYSYEV 120
50 TFDNGKALKY VQEKDLVTK RSKYSKGQIV KLKSSATADL DGSSLTDYRA SAGKIDHISY 180
NHSNTTGGYK YDITFDEGGK VTNIQEKDLD KVEVQLKSE NTAAQNNEIL KQAFAYAKQH 240
SGTILSLPNG EFKIGSQTPD KDYITLTS DT EIRGDNTTLL VEGSAYWFAF ATGTSASDGV 300
KNFTMRNINI KASDLEKGNQ FMIMADHGDN WKICNNSFTM VHKKGSHIFD LGSLQNSAFE 360
GNQFTGYAPE LTNVSKIDDN ADLHDFYSEV IQLDAAESSG VWDGGLIKAI DPNYENYNKE 420
55 KQLCNNITIA NNSFVPYIDS HGKIIAYSGT IGQHSSDVGL VKIYDNVFSN SLVSRFNQNG 480
KSEAWIFKAI HLKSNYNNAV YANSIS 506
<212> Type : PRT
<211> Length : 506
SequenceName : SEQ ID 357
60 SequenceDescription :

65
Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MRKLKVALFA SSILGMLAVS SYTAADTEDN QVTISHYNEQ AGTFDVNAVQ AANGKTIQSI 60
DVAIWSEENG QDDLKWHAS NDGSNQLTVH FNAENHGSKV GSYIAHAYIT YTDGNRVGVN 120

LGKRKLSLSA PQLSLKQGGL QLFSKLKPSA ADQLFSAVWS DENGQDDLHW YTADADGNTL 180
AGYANHKGYG TYHVHTYLKQ NGKMIPISAQ DIDIPKPKVK IQIDKINDTS YDVVVNNVPP 240
YISSVAIPVW SEQNGQDDLK WYQATKVADG IFKTTVYLKT HRFELGNYQA HIYGDSQLSK 300
KLDGLGETHF NVPSIINYED PQVTIDHYN I NKGTFDVTVA ETDNSKAIQS ISAAVWSDAN 360
5 QANLYWYEAK QLANGKAAIT VDVQKHGNT Q GSYNVHVYVH YNDGTTSGHV LANQQLNQIV 420
HYQPSAVRIT AYMNEKNTYP VGQCTWGVKE LAPWIPNWLG NGGQWASTVA VKGFKIGTVP 480
KVGAIACWSD GGYGHVAVVT HVESNNRIQV KEANYKNQQY ISNFRGWFD P TTSYLGRLTY 540
IYPD 544
<212> Type : PRT
10 <211> Length : 544
SequenceName : SEQ ID 358
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MANNYSRRQQ PTKKTKGTSR KRPTEHIKTG FSALQKSVAI IAGILGIITA LITINNYRNS 60
SHNDKKDSTS KTTIIKEKEV DDSNSNNNAA NSQAENDSNN NNNSAESNQ N QTATTANDSN 120
20 SNSANQNQAN SQSQANNQON QNNANAGQ 148
<212> Type : PRT
<211> Length : 148
SequenceName : SEQ ID 359
SequenceDescription :

25 Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
30 MKIFSFGTIR NNTALKPNYD DTTAFSGFGT IRNNTALKQS TNCASWFNRF GTIRNNTALK 60
LTIILINGVSF CFGTIRNNTA LKPRGPIFVS TFRNRAIHL S QISASK 106
<212> Type : PRT
<211> Length : 106
SequenceName : SEQ ID 360
35 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
40 <400> PreSequenceString :
MKRKRNL YFL IGLFLT VFL I IGCSMQK KTK SESSSTSQKT TLQTKQSSEK STDAKQTTEA 60
HSESSQSSSH SNNEETLAPI DTGAVLKADY SSMAGTWKNE EGQTLTFDQR GLTTPGMTVS 120
LLNIDQDGNL LLNVETGTTK NLTLYIVPAN KTLNQNQYFSN QSDSDSKTK DRIVSSESLN 180
SGKFTNRVYY HVSTH 195
45 <212> Type : PRT
<211> Length : 195
SequenceName : SEQ ID 361
SequenceDescription :

50 Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
55 MTPKKIKIAL TALISLMLAL FLFLFNHHSV RENSQQEKLK ISKASSKKSQ TSTSSVMTSS 60
RKATEQTSQA QTQSQSQA EQ SNPENVILPIP QELVGTYKGS SPQASEITFT ISSNGQLRAQ 120
ANFD PASDIN DVTATVSGVR KVGADTYIWE FVSGSSAALL PGVTGIGGLG KMQPGFILKG 180
GQLTPIMFTG SVDGEIDYSH PNPYPVSLNK Q 211
<212> Type : PRT
<211> Length : 211
60 SequenceName : SEQ ID 362
SequenceDescription :

Sequence

65 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKIINVIVL SLSVFFLIAC SNSSTGEKTS QSSEETKVRL IVKTDSNKTD EKVAFKKGAT 60

VMDVLKDNKYK VKESGGFITT IDGVTQDKKA GRYWMFDVND KLASKAADKI KVKNGDKIEF 120
YLKVKYKGKN 129
<212> Type : PRT
<211> Length : 129
5 SequenceName : SEQ ID 363
 SequenceDescription :

Sequence

10 <213> OrganismName : Streptococcus mutans UA159
 <400> PreSequenceString :
 MSNKPWEEKV TDATTDNEEM TRNSKDASII STPILTILLS LFFLIIGIL FFFVLYTSNGG 60
 SNEKAATSGF YSSSKTVKKA KNEANSQTDE QTTEAETSSS ETTSSSSSDSD GETITVQGGE 120
 GAAAIAARAG ISVDKLYELN PEHMTGHWY ANPGDNIIK 160
15 <212> Type : PRT
 <211> Length : 160
 SequenceName : SEQ ID 364
 SequenceDescription :

20 Sequence

 <213> OrganismName : Streptococcus mutans UA159
 <400> PreSequenceString :
 MPDNRMNYSI DSNMQFPLVE ITLETGEFAY IQRGSMVYHT PSVTLNNTKVN GRGSGLGKLV 60
25 GAIGRSVTSG ESFFITQAVS NASDGKLALA PSMFGQVIAL ELGEKQYRLN DGAFLALDGS 120
 AQYQMQAQS SVGRALFGGQGG LFVMTTEGQG TLLANSFGSI KKIELQNQEI TIDNAHVVAW 180
 SRDLNYDIHL ENGFMQSIGT GEGVVNTFRG TGEIYVQSLN LQQFAGVLQG FITNTNR 237

 <212> Type : PRT
30 <211> Length : 237
 SequenceName : SEQ ID 365
 SequenceDescription :

Sequence

35 <213> OrganismName : Streptococcus mutans UA159
 <400> PreSequenceString :
 MKKNYFWYGL LGLLALYLIT IAFIPGFHIF FSNMLMLALF FMLIALSNRS IFFFFLALGF 60
 LSIYLDIFH FDYSTGPLFT GIIIGVILN SFLKPHYSYS YKGNHYFNMK QHANYIDNET 120
40 DVFLKTLFSE NTSYVTSQEL NKIIIDTKFG EQSVDLSQAQ FMTDSPEIHI DVSFGETNLR 180
 IPNNWKIINK THSPFASISF SGFPSTNGDF INVTLTGTVA MGSLNIQY 228
 <212> Type : PRT
 <211> Length : 228
 SequenceName : SEQ ID 366
45 SequenceDescription :

Sequence

50 <213> OrganismName : Streptococcus pneumoniae R6
 <400> PreSequenceString :
 MKSITKKIKA TLAGVAALFA VFAPSFVSAQ ESSTYTVKEG DTLSEIAETH NTTVEKLAEN 60
 NHIDNIHLIY VDQELVIDGP VAPVATPAPA TYAAPAAQDE TVSAPVAETP VVSETVVSTV 120
 SGSEAEAKEW IAQKESGGSY TATNGRYIGR YGSWTAAKNF WLNNGWY 167
 <212> Type : PRT
55 <211> Length : 167
 SequenceName : SEQ ID 367
 SequenceDescription :

Sequence

60 <213> OrganismName : Streptococcus pneumoniae R6
 <400> PreSequenceString :
 MKHSHKKSFD WYSMQQRYSI RKYFPGAASV LLGTALVLGA AASVQTVQAE ENKQETTNSI 60
 SVGRGEAATK PAEVSASNKE KTYAAPTVAN PVETTPVKTE EVTKPAEKVE EAKDKKEEVT 120
65 HQDAVDKSKL LTALSRAKKL ESKLYTEASA ANLQTSIQAG QSLGKADAT EAELSAAESS 180
 IQSFIIGLEL RSNSNKETVS ETPVAKKADA VESKEGAKPA ATTERSADVS AILPTSTADK 240
 VETTSAPASI NEILKLGLSL SDARQNPAIR KEDVNRGYSG FRAASNPANP IVSGSGNTVA 300

| | | | | | | | |
|----|--|------------|------------|------------|------------|------------|------|
| | FADISQGGRS | YSFRGYGNSR | GGNSIHYDVT | TVRSGNSVNF | TISYSAPGDS | REFVNNNFIL | 360 |
| | DKGDGFGNPS | NATITSSNPR | VREQSKSISQ | GANYVSHSGY | SMTSAISTNT | EQTIRFSLPI | 420 |
| | INLNGDLSVR | LKPVTFNVDQ | GGGGAATSND | PYSNSNYYYR | ANPLYLDANP | YGGTNNKTVS | 480 |
| | EDIDFQTVYL | PTSKLPEGQT | RLVREGEKGQ | RQITYKVHFR | GNETLLGLPI | SNSVTKEAKP | 540 |
| 5 | RIMQIGVAKD | LIDTVKPRVD | QNKVGDTNNL | TFYLDNDGNG | VYTEGVDELV | QKIAIKDGAK | 600 |
| | GEKGDQGERG | LTGAKGEKGD | RGERGLTGAQ | GAKGEKGRG | ERGLTGAQGA | KGEKGRGER | 660 |
| | GLTGAQGAKE | EKGDRGERGL | TGAQGAKEK | GAQGERGLTG | AQGAKEKGD | QGERGLTGAQ | 720 |
| | GAKGEKGDQG | ERGLTGAQGA | KGEKGAQGER | GLTGTQGAKE | EKGDRGERGL | TGAQGAKEK | 780 |
| | GDRGERGLTG | AQGAKEKGA | QGERGLTGAQ | GAKGEKGDQG | ERGLTGAQGE | KGDRGERGLT | 840 |
| 10 | GAKGEKGDQG | ERGITGAKE | KGAQGERGLT | GAQGAKEK | DQGERGLTGA | QGEKGAQGA | 900 |
| | GRDGVTPTVT | VKDNKNDGTH | TITINDGRGN | VTSTVVRDGF | DGASPLVATQ | RNDADKTTTV | 960 |
| | IFYDKNGNN | ELDASDKKLK | EVVIADGAKG | EKGDKGEQGL | QGRDGEQGP | GEDGKTPTVK | 1020 |
| | VTDGQDGTHT | ITINDGKGGI | TTTVVRDGF | GASPLVSTHR | NEADKTTTVI | FYYDLNDNNQ | 1080 |
| | FDEGDTKLKE | VVIADGKQGP | KGDKGDNKGF | GFTPEVTVTD | NNNGTHTITI | TQPDNRPSLT | 1140 |
| 15 | TIVKNGEDGK | TPKVKAERDD | AKKQTTTLTY | IDKDGDSYT | AGKDELVTQT | VVKDGDGAA | 1200 |
| | GASGRDGKEV | LNGKVDPTTE | GKGDGTFVNT | QTGDVFKKG | NTWEPAGNIK | GPKGDKGADG | 1260 |
| | AKGEKGAQGE | RGLTGAQGVK | GEKGDQGERG | LTGSKGEKGD | QGERGLTGAQ | GAKGDKGEQG | 1320 |
| | LQGRDGAQGP | KGADGQGP | GPQGPKEQG | NPCTPGKDGK | SLIAVKNGVL | VTITPVEGRP | 1380 |
| | QTTTFVEDGQK | GADGKTPTVT | ITEGQNGTHT | LTVHNPSPD | VTTTIRDGAT | GQAGRDGKDV | 1440 |
| 20 | LNGKVNPPQN | QKNGDKYIN | IETGDVYVKN | NGNWDKEGNI | KGPKGDKGAD | GAKGEKGDQG | 1500 |
| | ERGLTGAQGA | KGADGAVGRD | GRDGKDVNLG | KANPEAHQK | DGDKYVNTET | GDVFKVNNGN | 1560 |
| | WDKEGNKGP | KGDKGADGAK | GEKGRGERG | LTGAQGAKE | DGAAGRDGRD | GRDGKDVNLG | 1620 |
| | KVNPEANQK | DGDKYVNTET | GDVFKVNNGN | WDKEGNKGS | KGDKGERGED | GKTPEVTVTP | 1680 |
| | GKDGHSTDT | FTVPGKDPVT | VNVKDGENGL | NGKTPKVDLL | RVQKNGNPS | HTIVTFYTDE | 1740 |
| 25 | NNDGKYTPGT | DELLGSEMIK | DGAKGADGRD | GKSLTVKDG | KETKVYQEDP | ANPGQPLNPE | 1800 |
| | KPLAVIRGV | DGKSPTVTAV | RKDEAGHKGV | EITVDNHDGS | QPTTVFVQDG | AKGKTGATGQ | 1860 |
| | DGQTPITTQ | RGQDQSTVV | TITTSKDPV | TFTVKDGKNG | KDGRAPKIKV | EDITSPSRIR | 1920 |
| | RDTDAATPT | RNGIRVTVD | DVNDNGVYDE | GVDKVLNSKD | IYNGIDGRDG | SAPTITTKDN | 1980 |
| | GDGHTITVQ | NPDGSESTTV | VKDGDGKTA | NITTENPDG | SHTITVTNPD | GSTKETVVKN | 2040 |
| 30 | GKDGKTPKVE | VTDNNDGHT | VKVTGDGDNV | TNAIKDGKD | GKAATATTTE | NPDGSHTVTI | 2100 |
| | TNPDGKNEF | VVKNGRDGVD | GRTPTASVRD | NGDGSHTIVI | TNPEGVTTET | TVRDGKSPKV | 2160 |
| | TITDEQNGTH | KISVLNGDGT | TTETIIKDGK | SPVATVRDNQ | DGTYTIRVEN | GNGTVSETTV | 2220 |
| | RDGKSPTAKV | VDNGDGHTI | TVNSDGITT | TTTVRDGREG | KLEVIDNNDG | SHTIKVTGAD | 2280 |
| | GKGTTTTIFD | GKSPKANIVD | NGDGHTTLTI | VDSGREYKS | IIKDGKDGK | SVSPTVTVKN | 2340 |
| 35 | NNDGTHVVTI | TNPDGSKTEM | VIKDGKDGKS | PKVSVEDNGD | GSHTITIINS | DGTVTKTVIK | 2400 |
| | DGKDGKGRD | GRDGKDGKDG | KCGCQDKPVT | PSNDKPVPPT | PNVPTPEVPV | KPVPAQPTPN | 2460 |
| | VPTPEVPVQP | TPAVSTPEVP | VKPVPAVPEQ | PVPTPAQPA | TPVNANPVAP | TTGKENRGDK | 2520 |
| | LPETGSQSDY | ISVLLGSGIL | LSLYVGRRKE | D | | | 2551 |
| | <212> Type : PRT | | | | | | |
| 40 | <211> Length : 2551 | | | | | | |
| | SequenceName : SEQ ID 368 | | | | | | |
| | SequenceDescription : | | | | | | |
| | Sequence | | | | | | |
| 45 | ----- | | | | | | |
| | <213> OrganismName : Streptococcus pneumoniae R6 | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | MKKRMLLAST | VALSFAPVLA | TQAEVLWTA | RSVEQIQNDL | TKTDNKTSYT | VOYGDTLSTI | 60 |
| | AEALGVDVT | LANLNKITNM | DLIFPETVLT | TTVNEAEVET | EVEIQTPQAD | SSEEVTTATA | 120 |
| 50 | DLTTNQVTV | DQTVQVADLS | QPIAEAPKEV | ASSSEVTKTV | IASEEVAPST | GTSVPPEQTA | 180 |
| | ETSSAVAEAA | PQETTPAEKQ | ETQTSPQAAS | AVEATTTSSE | AKEVASSNGA | TAAVSTYQPE | 240 |
| | ETKIISTTYE | APAAPDYAGL | AVAKSENAGL | QPQTAAFKEE | IANLFGITSF | SGYRPGDSGD | 300 |
| | HGKGLAIDFM | VPERSELGDK | IAEYAIQNMA | SRGISYIIWK | QRFYAPFDSK | YGPANTWNPM | 360 |
| | PDRGSVTENH | YDHVHVSMMG | | | | | 380 |
| 55 | <212> Type : PRT | | | | | | |
| | <211> Length : 380 | | | | | | |
| | SequenceName : SEQ ID 369 | | | | | | |
| | SequenceDescription : | | | | | | |
| | Sequence | | | | | | |
| 60 | ----- | | | | | | |
| | <213> OrganismName : Streptococcus pneumoniae R6 | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | MTILGKDTVQ | QSAKGESVTQ | EATPEYKLEN | TPGGDKGGNT | GSSDANANEG | GGSQAGGSAH | 60 |
| 65 | TGSQNSAQSQ | ASKQLATEKE | SAKNAIEKAA | KNKQDEIKGA | PLSDKEKAEL | LARVEAEKQA | 120 |
| | ALKEIENAKT | MEDVKEAETI | GVQAIAMVTV | PKRPVAPNAA | PKTTSAPQAT | AGTMQDVTYQ | 180 |
| | SPAGKQLPNT | GSASSAALAS | LGLVVATSGF | ALLGRKTRRR | K | | 221 |

<212> Type : PRT
<211> Length : 221
SequenceName : SEQ ID 370
SequenceDescription :

5
Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :

10 MMTTGCSMGA YHALNFFLQH PDVFTKVIAL SGVYDARFFV GDYYNDDAIY QNSPVDYIWN 60
QNDGWFIDRY RQAEIVLCTG LGAWEQDGLP SFYKLKEAFD KKQIPAWFAE WGHDDVAHDWE 120
WWRKQMPYFL GNLYL 135
<212> Type : PRT
<211> Length : 135

15 SequenceName : SEQ ID 371
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :

20 MNKGLFEKRC KYSIRKFSLG VASVMIGATF FGTSPVLADS VQSGSTANLP ADLATALATA 60
KENDGHDFEA PKVGEDQGSP EVTDGPKTEE ELLALEKEKP AEEKPKEDKP AAKPETPKT 120
VTPWQTVTEK KEQQGTVTIR EEKGVRYNQL SSTAQNDNAG KPALFEKKGL TVDANGNATV 180
25 DLTFKDDSEK GKSFRGVFLK FKDTKNNVFFV GYDKDGFWE YKSPTTSTWY RGSRVAAAPET 240
GSTNRLSITL KSDGQLNASN NDVNLFDTVT LPAAVNDHLK NEKKILLKAG SYDDERTVVS 300
VKTDNQEGVK TEDTPAETK GPEVDDSKVT YDTIQSKVLK AVIDQAFPRV KEYSLNHHTL 360
PGQVQQFNQV FINNHRTPE VTYKKINETT AEYLMKLRDD AHLINAEMTV RLQVVDNQLH 420
FDVTKIVNHN QVTPGQKIDD ERKLLSSISF LGNALVSVSS DQTGAKFDGA TMSNNTHVSG 480
30 DDHIDVTNPM KDLAKGYMYG FVSTDKLAAG VWSNSQNSYG GGSNDWTRLT AYKETVGNAN 540
YVGIHSSEWQ WEKAYKGIVF PEYTKELPSA KVVITEDANA DKKVDWQDGA IAYRSIMNNP 600
QGWWKVKDIT AYRIAMNFGS QAQNPFMLTL DGIKKINLHT DGLGQGVLLK GYGSEGHDSG 660
HLNYADIGKR IGGVEDFKTL IEKAKKYGAH LGIHVNASET YPESKYFNEK ILRKNPDGSY 720
SYGWNWLDQG INIDAAYDLA HGRLARWEDL KKKLGDGLDF IYVDVWNGNQ SGDNGAWATH 780
35 VLAKEINKQG WRFAIEWGHG GEYDSTFHHW AADLTGGYT NKGINSATR FIRNHQKDAW 840
VGDRSYGGA ANYPLLGGYS MKDFEGWQGR SDYNGYVTNL FAHDVMTKYF QHFTVSKWEN 900
GTPVTMTDNG STYKWTPEMR VELVDADNNK VVVTRKSNDV NSPQYRERTV TLNGRVIQDG 960
SAYLTPWNWD ANGKKLSTDK EKMYFNTQA GATTWTLPSD WAKSKVLYK LTDQKTEEQ 1020
ELTVKDGKIT LDLLANQPYV LYRSKQTNPE MSWSEGMIY DQGFNSGTLK HWTISGDASK 1080
40 AEIVKSQGAN DMLRIQGNKE KVSLTQKLTG LKPNTKYAVY VGVDNRSNAK ASITVNTGEK 1140
EVTTYTNKSL ALNYVKAYAH NTRRNATVD DTSYFQNMVA FFTTGSVDVN VTLTSLREAG 1200
DEATYFDEIR TFENNSSMYG DKHDTGKGTG KQDFENVAQG IFPFVVGVE GVEDNRTHLS 1260
EKHDPYTQRG WNGKKVDDVI EGNWSLKTNG LVSRRLVYQ TIPQNFREFA GKTYRVTFEY 1320
EAGSDNTYAF VVGKGEFQSG RRGTAQSNLE MHELPNTWTD SKKAKKATFL VTGAETGDTW 1380
45 VGIYSTGNAS NTRGDSGGNA NFRGYNDFMM DNLQIEEITL TGKMLTENAL KNYLPTVAMT 1440
NYTKESMDAL KEAVFNLSQA DDDISVEEAR AEIAKIEALK NALVQKKTAL VADDFASLTA 1500
PAQAQEGLAN AFDGNLSSLW HTSWGCGDVG KPATMVLKEA TEITGLRYVP RSGSGNGNLR 1560
DVKLVTDES GKEHTFTATD WPDNNKPKDI DFGKTIKAKK IVLTGKTGYG DGGDKYQSAA 1620
ELIFTRPQVA ETPLDLSGYE AALAKAQKLT DKDNQEEVAS VQASMKYATD NHLLTERMVE 1680
50 YFADYLNQLK DSATKPDAPT VEKPEFKLSS VASDQGKTPD YKQEIARPET PEQILPATGE 1740
SQFDTALFLA SVSLALSALF VVKTKKD 1767
<212> Type : PRT
<211> Length : 1767
SequenceName : SEQ ID 372
SequenceDescription :

55
Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :

60 MKLYNKSELR YSRIFFDKRP PAFAFILIIS TAILSGALV GAAYIPKNYI VKANGNSVIT 60
GTEFLSAISS GKVVTLHKSE GDMVNAGDVI ISLSSGQEGE QASSLNKQLV KLRAKEAIFQ 120
KFEQSLNEKY NRMSNSGEEQ EYYGKVEYYL SQLNSENYNN GTQYSKIQDE YTKLNKITAE 180
RNQLDADLQT LQNELIQLQQ QGDSPSLSDT TSADDAKLE TKILEITTKI EALKTNITSK 240
65 NSEIDSQQSN IKDMNRTYND PTSQAYNIYA QLVSELGTAR SNNNKSITEL EANLGVATGQ 300
DKAHSILAPN EGTLLHYLVPL KQGMSIQQGQ TIAEVSGKEK GYYVEAFVLA SDISRVSKGA 360
KVDVAITGVN SQKYGTLLKGQ VRQIDSGTIS QETKEGNISL YKVMIELETL TLKHGSETVV 420

LQKDMPEVEVR IVYDKETYLD WILEMLSFQK 450
<212> Type : PRT
<211> Length : 450
SequenceName : SEQ ID 373
SequenceDescription :

5

Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :

10 MNKGLHRIIF SKKHSTMVAV AETANSQGKG KQAGSSVSVS LKTSGDLCGK LKTTLKTLVC 60
SLVSLSMVLP AHAQITTDKS APKNQQVVIL KTNTGAPLVN IQTPNGRGLS HNRYTQFDVD 120
NKGAVLNNDR NNNPFLVKGS AQLILNEVRG TASKLNGIVT VGGQKADVII ANPNGITVNG 180
GGFKNVGRGI LTIGAPQIGK DGALTGFQDVR QGTLTVGAAG WNDKGGADYT GVLARAVALQ 240
15 GKLGQKNLAV STGPQKVDYA SGEISAGTAA GTKPTIALDT AALGGMVADS ITLIANEKGV 300
GVKNAGTLEA AKQLIVTSSG RIENSGRIAT TADGTEASPT YLSIETTEKG AAGTFISNGG 360
RIESKGLLVI ETGEDISLRN GAVVQNNNGSR PATTVLNAGH NLVIESKTNV NNAKGSANLS 420
AGGRTTINDA TIQAGSSVYS STKGDELGE NTRIIAENVV VLSNGSIGSA AVIEAKDTAH 480
IESGKPLSLE TSTVASNIRL NNGNIKGGKQ LALLADDNIT AKTTNLNTPG NLYVHTGKDL 540
20 NLNVDKDLSA ASIHLSKDNA AHITGTSKTL TASKDMGVEA GLLNVTNTNL RTNSGNLHIQ 600
AAKGNILRN TKLNAKALE TTALQGNIVS DGLHAVSADG HVSLLANGNA DFTGHNTLTA 660
KADVNAAGSVG KGRKADNTN ITSSSGDITL VAGNGIQLGD GKQRNSINGK HISIKNNGGN 720
ADLKNLNVHA KSGALNIHSD RALSIENTKL ESTHNTLNA QHERVTLNQV DAYAHRHLSI 780
TGSQIWQNDK LPSANKLVAN GVLALNARYS QIADNTTLRA GAINLTAGTA LVKRGNIWNS 840
25 TVSTKTLEDN AELKPLAGRL NIEAGSGTTL IEPANRISAH TDLSTKTGGK LLLSAKGGNA 900
GAPSAQVSSL EAKGNIRLVT GETDLRGSKI TAGKNLVVAT TKGKLNIEAV NNSFSNYFPT 960
QKAAELNOKS KELEQQIAQL KKSSPKSKLI PTLQEEERDL AFYIQAINKE VKGKKPKGKE 1020
YLQAKLSAQN IDLISAQIE ISGSDITASK KLNLAHAAGVL PKAADSEAAA ILIDGITDQY 1080
EIGKPTYKSH YDKAALNKPS RLTGRTGVSI HAAALDDAR IIGASEIKA PSGSIDIKAH 1140
30 SDIVLEAGQN DAYTFLKTKG KSGKIIRKTK FTSTRDHLIM PAPVELTANG ITLQAGGNIE 1200
ANTTRFNAPA GKVTLVAGEE LQLLAEEGHI KHELDVQKSR RFIGIKVGKS NYSKNELNET 1260
KLPVRVVAQT AATRSQWDTV LEGTEFKTTL AGADIQAGVG EKARVDKII LKGIVNRIQS 1320
EEKLETNSTV WQKQAGRGST IETLKLPSFE SPTPPKLSAP GGYIVDIPKG NLKTEIEKLS 1380
KQPEYAYLKQ LQVAKNINWN QVQLAYDRWD YKQEGLTEAG AAIILAVTV VTSGAGTGAV 1440
35 LGLNGAAAAA TDAAFASLAS QASVSFINNK GDVGKTLKEL GRSSTVKNLV VAAATAGVAD 1500
KIGASALNNV SDKQWINNLT VNLANAGSAA LINTAINGGS LKDNLGDAAL GAIIVSTVHGE 1560
VASKIKFNLS EDYITHKIAH AIAGCAAAAA NKGKQDGA IGAAGVEIVGE ALTNGKNPAT 1620
LTAKEREQIL AYSKLVAGTV SGVVGQDVNT AANAAKVAIE NNLSQEEYA LREKLIKAK 1680
GKGLSLDWG SLTEQEARQF IYLIKEDRYN NQLLDYQKN PSSLNQEKI ILAYFINQTS 1740
40 GGNTAWAASI LKTPQSMGNL TIPSKDINNT LSKAYQTLR YDSFDYKSAV AAQPALYLLN 1800
GPLGFSVAAA TVAAGGYNIG QGAKAISNGE YLHGTQVNV GTLMVAGSVS AQAAISAKPA 1860
PVTRYLSNDS APALRQALTA ESQRIRMKLP EEYRQIGNLA IAKIDVKGLP QRMEAFSSFQ 1920
KGEHGFISLP ETKIFKPISV DKYHNIAAPP RGTLRNIDG YKLLLETIAQQ LGNNRNVSGR 1980
IDLFTELKAC QSCSNVILEF RNRYPNIQLN IFTGK 2015

45 <212> Type : PRT
<211> Length : 2015
SequenceName : SEQ ID 374
SequenceDescription :

50

Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :

55 MDLIQTPNKQ FVDGDRRTPG TPVPAWNLN LQGEYSILN AVGIEPNKAD HAQVLSAIKT 60
LAADASQVAS IDALRKYSGT GYVNVNAYHA NTTVGGGVFV ADKADKSTAD NGCTVIVSTD 120
GTRWKRVSFG MLNLHDFGYV ASKNNALSTL NAAESAALDV VVDCLGLSID TGNIPQKKNK 180
YTNGKFVING KTVQVQYQPI RSGIGRFISG TGAAANLKS EWTGAGLIVI GEGAMEQMEK 240
CVSSIAIGDR AQGFSKVS RD NIAIGADSLI NVQAATEWYD QSRMEGTRNI GIGGNAGRGI 300
TSGYSNVSIG RNAGQGLGEG SSNIALGAGA MAGTAPVGFS GDIEVFWPSS TSRTIAIGEA 360
60 VLQTYQGRAA QTAIGANAAR NTKKAKEVTA IGSAAMENLE RNRAPNGGDV VWTGTEAGTY 420
AQSGKNITLT FPNIRGAQAT YWVGIRLTSG TAQTLQNDVV PAQVVSVMGN TLIIQSSKEL 480
TATGAAELKY VYSVNSTATK NEELTIIGAN AMNKALTAGY STIIGVDAAL LGDNYQKTTA 540
IGASSLRTGS HISTTAIGYW VIPLASSEKC VAIGDSAGYR NVQGDFTLTK ITNSIAIGYG 600
ARINGDNEIQ IGTGQTLTYA PTAVNIRSDG RDKADVPLT NGLDFVMKLG PMTGYDTRD 660
65 SYVDELFDL PADERADKVR EWWANPIKDG SHKEDRLRH FIAQDIAALE DEYGRLLPMVN 720
KTNDTYTVEY ETFIPVLTKA IQEMAARIET LETEMKESKK 760

<212> Type : PRT

<211> Length : 760
SequenceName : SEQ ID 375
SequenceDescription :

5 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
10 MKNISRKCFM TSVVCIILGG ILLGAGYATG GLQDIKHQTA PKKVIKTFDQ ITALDIDSSA 60
STITVETGPV QRPTVTYYTH PKFIDPIVTT LTGKTLSSLSQ KPKDIVITGG IEILGFTLNN 120
SRQEKNYRSI TITVPEKTSI NEVKGSNVPH TTLSNLTVDQ MQFDGNLTLL HTKVKKATIT 180
GMLEATKSQL TNLELKADYS FSNLTDSSVE NGTISLGNGQ LTTKDTTLKA INIQSLHPGG 240
IEAERTLEN VTFTVSKSKE EEEENDYYDN DAIFTAHALT LKGTNTISGG DIDVDITLTK 300
AKAIAYRART ENGKVSLSQS LTPAKIGKES TSDVISYVAE NKAATGNLTV NLNKGDITIK 360

<212> Type : PRT
<211> Length : 360
SequenceName : SEQ ID 376
SequenceDescription :

20 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
25 MFKKENLKQR YFNFGI VALA LTILAIIFAF SSKNADTKSY AKKSESKMVT IDKAPKNNHA 60
ITKEESKEKA KSIASEPIPT VENSVAPTVT EEAPVVQQEV TQTVQQVSSV AYNPNNVVLS 120
NGNTAGIVGS QAAAQMAAAT GVPQSTWEHI IARESNNGNPN AANASGASGL FQTMPGWGST 180
ATVEDQVNAA LKAYSAGGLS AWGY 204

<212> Type : PRT
30 <211> Length : 204
SequenceName : SEQ ID 377
SequenceDescription :

35 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
40 MLEELKTLIK NPKLMITMIG VALVPALYNL SFLGSMWDPY GRVNDLPPIAV VNHDKPAKRA 60
DKSLTIGNDM VDKMSKSKDL DYHFVSSKSA QKGLKKGDYY MVITLPEDLS QRATTLLNPE 120
PQKLITIRYQT SKGHGMVAAK MGETAMAKLK ESVSQNIKT YTSAVFSSMT DLQSGLEKAS 180
TGSQALDSGA KTAQMGSQL SDNLAGLSSA SWQFQQGTNR LTSGLTAYTA GVSQVKDGLG 240
QLSTDMPVYL NGVSRLSQGA SQLNQGLSQL TQSTTLSDDK AKRIQSLEVG LPVLNQGIQQ 300
LNENLSTMVQ PKLNTDELGN NLAAIAQAAQ QLLVKEAAAH KEQLAVLQAT SAYQSLTAEQ 360
QGELTAALTQ TDKGEAVAPA QTILRSVQTL STSLQSLSQE DQSKQLEQLK EAVAQIANQS 420
45 NQALPGASSA LTELSTGLAK VNGSLNQQVL PGSNQLTTGL AQLNRYNTAI GSGVIKLSEG 480
ANALSSKSSE LLDGSHQLSE GATKLADGSS QLSQGGHQLT SGLTELSTGL SILNGSLAKA 540
SQQLSLVSVT DKNKAVAKP LVLNEKDKDG VKTNGIGMAP YMIAVSLMVV ALSTNVIFAN 600
SLSGRPVKDK WDWAQKQFVI NGFISTMGSI VLYLAIQLG FEARYGMETL GFIMLSGWTF 660
MALVTALVGW DDRYGSFASL VMLLLQVGSS GGSYPLELSG AFFQKLHPFL PMTYVVVSGLR 720
50 QTISLSGHIG VEVKVLGTGL LAFMVLSELLI YRPKKTV 757

<212> Type : PRT
<211> Length : 757
SequenceName : SEQ ID 378
SequenceDescription :

55 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
60 MSRDPTYTIN EHDLSFADGR FYVTFKADKS SETVRLNSSC LGNTIIKKLQ VEDDNTMHDF 60
VKPKVTTQQA FGLAQVVKEL DLQLKDPKSD LWGKIKFNNK AMLVEYANKE MSSAIAQSAE 120
QILLQVKSID DERYSKFEQT LNGIKQTVKS ESVESARTQL ASMFDSRISG LDGKYSRLSQ 180
TIDSLSSRLD DGVGNYSLS QKVSGIDLRV SNAANDVSRL SQTAQGLQSQ ITNANQNYSS 240
LSQTVQGLQT TVRDNQSNAT SRINQLSDLI STKVSKGDVE TTIAQSYDKI AFAIRDKLPA 300
65 SKMSGSEIIS AINLDRSGVK ITGKNITLDG NSYISNAVIK DAHIANMDAG KINTGYLNAN 360
RIATEAITGE KIKMDYAFFN KLTANEGYFR TLFADIFAT SVQSVTLAS KITGGVLAAT 420
NGASQWDLNN ANMTFNRDAT INFNSKNNAL VRKDGHTAF VHFSNATPKG YRGSALYASI 480

GITSSGDGID SASSGRFAGL RSFRYATGYN HTAAVDQTEL YGDNVLIADD FSINRGFKFR 540
PDKMEKVLDL NDLVAAVVAL GRCWGHLANV GWNTAHSNFT SAVSRELNNY ITKI 594

5 <212> Type : PRT
<211> Length : 594
SequenceName : SEQ ID 379
SequenceDescription :

10 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MAADGKVTIL VDVDGKQVKV LNSELDKVAK HGDKGSSSLK KFAVGAGVFK LASAAVDLVS 60
QSLGKAITRE DTLEKYPRVM KAMGHAEDV ARSTDKIANG IDGLPTTLDE VVGTAQRLTS 120
15 ITKDINKSTN LTLALNNAFL ASGASSEAS RGLEQYAQML SAGKVDMQAW KTLQETMPYA 180
LQQTAEAFGF AGASAQKDFY EALKNGQITF DQFSNKLIEL NDGVGGFAEL AKENSKGIET 240
SFNNIKNAIA KGVANSIKAL DDLSKAATGK GIADHFDSLK VVINASFSAI NASIKASTPL 300
FKLLFSVIGA GISVVKALSP ALVGVASGLA AMRAVNETIT MIKALNRAWV MASASMSIGA 360
TTIKTVTAVQ AVSTTMTKAD MVARLSQLGV LKASTVIYGV MTGAISLSTA ATIASTAAVT 420
20 ALKAALVALT GPVGWVVGAI GALVAVGVSL WSWLTKESE TKKLKKEQEG LVESNKQLRD 480
SVREGVQERK KGLESVKEST AAHQKLADEI IKLAAKENKT AGEKQNLKNK IDQLNGSIDG 540
LNLAYDKNSN SLSHNADQIK SRISAMEAES TWQTAQQNLL NIEQKRSEVS KKLAENADLR 600
KKWNEEANVS DSVRKEKIAE LTEEEAKLKN MQTQLQEEYN KTSATQQAAA DAMAAAEESG 660
SARQVIAYEN MSEAQRTAID NMRTKYSLL ETTTSIFDAI EQKTALSVDQ MNTNLEKNRA 720
25 ATEQWATNLE ILAQRGVDQG ILEQLRRMGF EGATQTQVFV DATDAELAPL QENFRAATET 780
AKNAMGSVLD SAGVEMPEKV KGMVTNVSTG LQAELOAANF AQLGQEIPNG VSQGISQGAG 840
KASDASVKMG QEVKRSFQGE LGIHSPSRVF TEYGGHITDG LSNQVTNGTS KVMQTMQSLA 900
QQMSQKGQOI VNDMRSSKNQ ITDAFSTMSG PMHSHGVNAM QGLANGIYAG SGAALAAAQS 960
IAARITATIQ SALDIHSPSR VMRDEVGRFI PQGIAVGIDA DRKVIDSSMQ KLKESMTINA 1020
30 TPEIASGFEG GVAGIANQTT NNSNNSFTLN VKVDESDGNS HEKYQRLFRE FSWYIQQQQG 1080
RLGDVK 1086
<212> Type : PRT
<211> Length : 1086
SequenceName : SEQ ID 380
35 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
40 <400> PreSequenceString :
MAKEPWEEKI VDDTIGTRTR KSRNAFISTP WLTALLSVFF VIIVAILFIF FYTSNSGNSR 60
QAETNGFYGA STHKKTRKAS NAKKTSSSST TDTTPSSEE TLASSEGTGE TLTVLAGEGA 120
ASIAARAGIS VEQLQALNPE HMTQGYWYAN PGDQVTIK 158
<212> Type : PRT
45 <211> Length : 158
SequenceName : SEQ ID 381
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MSKRGKIKIT TKTKLITASV ITLVLIITGV VLWKQQQNTL TADIAKEPYS TVSVTEGSIA 60
SSTLLSGTVK ALSEYIYFD ANKGNDATVT VKIGDQVTQG QQLVQYNTTT AQSAYDTAVR 120
55 SLNKIGRQIN HLKTYGVPV STETNKDEAT GEETTTTVQP SAQONANYKQ QLQDLNDAYA 180
DAQAEVNKAQ IALNDTVVIS SVSGTVVEVN NDIDPSSKNS QTLVHVATEG QLQVKGTLTE 240
YDLANVKVGQ SVKIKSKVYS NQEWTKISY VSNYPTESNA GSTTPAGSTG AGSSTGAAYD 300
YKIDIISPLN QLKQGFVSV EVVNEAKQAL VPLTAVIKD KKHYVWTYDD ATGKAKKVEV 360
TLGNADAQQQ EIHKGVAVG D IVIANPDKNI KPDKKLEGVI SIGTNTKPEK DSQSKNKKSG 420
60 VDK 423
<212> Type : PRT
<211> Length : 423
SequenceName : SEQ ID 382
SequenceDescription :

65 Sequence

<213> OrganismName : Treponema pallidum
<400> PreSequenceString :
MLRLPTARAC ITMGTMIRHT FTHRCGALLC ALALGSSTMA ATAAAKPKKG QMQKLRQRPV 60
WAPTGGRYAS LDGAFTALAN DASFFEANPA GSANMTHGEL AFFHTTGFGS FHAETLSYVG 120
5 QSGNWGYGAS MRMFFPESGF DFSTTTEPVC TPASNPIKQR GAIGIINFAR RIGGLSLGAN 180
LKAGFRDAQG LQHTSVSSDI GLQWVGNVAK SFTSEEPNLY IGLAATNLGL TVKVSDKIEN 240
CTSTCEKCGC CKERCCCNGK KACCKDCDCN CPCQDCNDKG TVHATDTMLR AGFAYRPFWS 300
FLFSLGATTS MNVQTLASSD AKSLYQNLAY SIGAMFDPFS FLSSLSSSFRI NHKANMRVGV 360
GAEARIARIK LNAGYRCDVS DISSGSGCTG AKASHYLSLG GAILLGRN 408
10 <212> Type : PRT
<211> Length : 408
SequenceName : SEQ ID 383
SequenceDescription :
15 Sequence

<213> OrganismName : Treponema pallidum
<400> PreSequenceString :
MSRTFRAWQC VGALCALSPL LPAYSSEGVR EVPPSQSPQV VVAYEPIRPG DQLLKIGIVA 60
20 GCQLYIAGGN GTNGSSSSGT NGNGNGKLLG GGGFHLGYEY FFTKNFSLGG QVSFECYRTT 120
GSNYYFSVPI TVNPTYTFAV GRWRIPLSLG VGLNIQSYLS KKAPGLIAEA SAGLYYQYTP 180
DWSIGGIVAY TQLGDIASSP DKCRAVGLAT IDFGVRYHF 219
<212> Type : PRT
<211> Length : 219
25 SequenceName : SEQ ID 384
SequenceDescription :
30 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgataaatt taagtaagga agcaacggtg gggaaagcat taacccttat tgctatactt 60
atgatgttgt cttttcttgt agcttctcaa gcggcgggat tagtcataaa aaatggaacg 120
35 gtatataacg ccaatggtgt gccagtcggt gacatcaaca aacctaacgg tagcggttta 180
tctcataata tctgggataa cctaaacggt gataaaaatg gtgtcgtttt caataatagc 240
gctaataaat ccagtacttc acttgccgga aatattcagg gaaacagtaa tctgacctcc 300
gggtcggcga aggtgatcct gaatgaggtt acttccaaaa atccttcaac cattaatggg 360
atgatggaag ttgcagggga taaagcggat ctgattattg ccaaccgaa tggattact 420
40 gtaaacggtg gcggttcaat caatacaggt aaacttacct taaccaccgg gacgcggat 480
atccaggatg acaagctggc cgggtactcc gtgaacggcg gtaccattac gctcggtaaa 540
ctggataacg ccagcccgac agaaattctg tcccgtaacg tggtagttaa cggcaaagt 600
tctgccgatg agctgaacgt tgttgctggc aataactatg ttaatgccgc aggccagggtg 660
accggtagcg tatccgccac ggggtccggt aacgggtaca gcgtagatgt tgccaaactg 720
45 ggcggaatgt atgcgaacaa aatcagtcgt gtgagcaccg agaaagggtg ggggggttcgc 780
aacctcggcg ttattgctgg ggggtgtaat ggtgtcagca tgcattccaa aggtaacctg 840
ttaaacagta acgcccagat tcagtcgtca agcagcatca acctgacaac aaatgggtact 900
ctggataaca ccaccggtac ggtgacatct gtaggcacta tctcgcttaa taccaacaag 960
aatactatcg tgaatacccc tgccggtaac atctctacga tgggcgatat ctacgttaac 1020
50 agcgggtacg ttgacaatac taacggcaag ctgcccgtg caggaatgct ggcgggtgat 1080
accaataacg ccacgctgat taactctggt aaagggagt ctgtcgggat tgaagcgggg 1140
ctcgtggcgc tgaaaaccgg aacgctcaac aacagcaatg gtcagattcg cgggtggctat 1200
gtgggtcttg aatccgctgc gctgaataac aacaacgggtg atatccagac caccggcgat 1260
atcgccatta tcagtaacgg taatgtggat aacaacaaag gtctgatccg ttcgtccacc 1320
55 gggcatatcg ttattggcgc ggcaggtagc gtaataatg gttcaaccaa aaccgccgat 1380
accggcagtt ctgactctct gggcattatt gcagataccg gcgtagaaat tgggtgcgaac 1440
aacatcaata acaacggcgg acagattgcg tctaattggca acgtctccct gtcaagttac 1500
agcacgatcg acgactatgc gggcaaaatt ctgtccaaca gcaaagtgat tatcaaggga 1560
agctctctgc gtaacgatac cggggggatc agcggtaagc aggtattga agtcgccgtt 1620
60 ggcggcagcc tgaccaataa tattggcgtg atcagctctg aagaggggtga tatctccctg 1680
ttagccaact ccgtggataa ccacggcggc ttcatgatgg ggcagaacat cacgatggag 1740
tcgatgtctg gcgtcaataa caacacagcg ctgatcgtgg ccagcaaaaa actgaagata 1800
aatgcgcgcg gcagtatcga aaaccgcat ggcaataact tcggtaatgc ttatgggtctg 1860
tacttcggca tgccctcagca aacgggtgga atgggtcggca aggaaggcat cgagctttcc 1920
65 gggcagaaca tctataacaa caacagccgt ctatatcgtg aggatgggtcc tctgactctg 1980
caggcgcaga acacgttcga caacacgcgt gctctggtca ccagcggggc ggatgcatct 2040
attcaggttg gcggaacgta ttacaacaac tacgctacca cctggagtgc gggcaacctg 2100

| | | | | | | | |
|----|---|-------------|-------------|-------------|-------------|-------------|------|
| 5 | gatatcgacg | cgaccacgct | gcaaaacagc | agcagcggta | cgatgatcga | taacaatgcg | 2160 |
| | accgggttca | tagcatctga | taaaaacctg | tcactggaag | tgggtgaatag | ccttaccaac | 2220 |
| | tacggctgga | tcagcggtaa | aggcgatgtt | gatgtcacgg | tgaataacgg | caacctgtat | 2280 |
| | aaccgcaata | ccattgcggc | tgaagggggg | ctggatattg | ccgcgttgaa | cggtattgaa | 2340 |
| | aactggaagg | atatttctgc | tggcggcgac | ctgacgatga | acaccaatcg | ccatgtgacc | 2400 |
| | aacaactcca | acagcaatat | ggtggggcag | aatattgtta | ttaacgcggg | taacgatatc | 2460 |
| | aacaaccgtg | gcaacattgt | cagtgcgct | gacctgaacg | tgacgaccaa | aggcaacctg | 2520 |
| | tataactatc | tctatatggg | agggtatggg | gatatcgcat | tgtcggcaaa | tagcgtggcg | 2580 |
| | aacaataacg | cgaccatcga | agcgacaggc | gatctgatta | tcgattcgaa | gggtaacgtg | 2640 |
| 10 | ggtaacaacc | gcggtaatct | gcatgcgttg | aacggcgtgt | tgtctgttaa | aggcaacaat | 2700 |
| | ctgaacaacg | ataacgggtg | aattcgtggg | tatggcgatg | tcacgctggc | actgacgggc | 2760 |
| | aactacgaca | gctataaggg | tctcgtgacc | tctgaaacgg | gcgacgtgac | tctgacggcg | 2820 |
| | aacattgtag | acaacgccta | tggtttgatt | gccgggtgaga | atgtttctgt | cgatgctaaa | 2880 |
| | tcgacgattt | acaacaacac | tgcgctgate | gcggcgaaat | aaaagctggg | tattaacgct | 2940 |
| 15 | ggcggcaacc | tcgaaaaccg | cgacgggaat | aacttcctgc | gtaataacgg | cgcgctgttt | 3000 |
| | ggaattaccg | acaacgttgg | cggcatcgta | ggtaaagaag | gtgtcacgct | ttctgctcag | 3060 |
| | aacgtctaca | acaataacag | cagcatcctc | gctgaaaatg | gtccgcttaa | tctgctgtcc | 3120 |
| | aggggaacgc | tggataatac | ccgcgcgctt | cttagcagtg | gggctgatgc | catcatccgt | 3180 |
| | gcggcagggg | cgttctacaa | caactatgcc | accacgtaca | gcgccggtaa | tctcgacgtt | 3240 |
| 20 | tatgcggcgt | cgttgaacaa | cgccagcgat | ggcgcctgg | aagacaatac | cgccacgggc | 3300 |
| | gtgattgcgt | ctgacaaaaa | cctggatctg | agcgttgata | acagtgtcac | taactatggg | 3360 |
| | tggatcagcg | gtaaaggaga | tgtgcatctt | aatgttctga | aaggcacgct | gtataaccgt | 3420 |
| | aatgccatcg | cggcggacaa | cgcgctgacc | attaatgccc | tgaacgggtg | tgagaacttt | 3480 |
| | aaagacattg | tggcgggtac | tgcgctgact | attgatacgc | agaagtatgt | taccaacaac | 3540 |
| 25 | agcaacagta | atatgttggg | acaaaccatc | gcgatcaatg | ccgtgaatga | cattaataac | 3600 |
| | cgtggaaata | ttgtgggtga | ttattctctg | gggtgttaaaa | ccaccggtaa | tatttataac | 3660 |
| | tacctcaata | tgctgagtta | tgggtgcgct | ggcgtatcgg | caaataaggt | tacgaatagc | 3720 |
| | ggtaaagacg | ctgttctcgg | tggcttctac | ggtttagcgt | tagaagcaaa | cgaaactgat | 3780 |
| | aacaccggta | ctattgtcgg | catgtaa | | | | 3807 |
| 30 | <212> Type : DNA | | | | | | |
| | <211> Length : 3807 | | | | | | |
| | SequenceName : SEQ ID 385 | | | | | | |
| | SequenceDescription : | | | | | | |
| 35 | Sequence | | | | | | |
| | ----- | | | | | | |
| | <213> OrganismName : Escherichia coli O157:H7 | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | gtgaacacaa | tacacttgcg | ctgtctcttc | aggatgaatc | ccctgggtctg | gtgcctgtgg | 60 |
| 40 | gctgatgttg | cagcaaagct | aaggctcgctt | aaacgctact | cagtattcac | ttttcagagg | 120 |
| | atgaaattta | tgaacaggac | cagtccttat | tattgtcgtc | gctcagtact | ttccttattg | 180 |
| | atatctgcct | tgatatatgc | ccgcgccggg | atggctgcct | tcactcctga | tgttattggg | 240 |
| | gtggtaaacg | atgagactgt | agatggcagc | caacgagtag | atgaacgagg | tacaacaaat | 300 |
| | aacactcata | ttatcaacca | tggccagcag | aatgtttatg | gcgggggtatc | taatggaagt | 360 |
| 45 | cttattgaat | ctgggtggata | tcaagatgta | ggaaggcata | acaattatgt | ggggcagtct | 420 |
| | aataatacca | ccattaacgg | gggcagacag | tcaattcatg | acgggggtat | ttccacaggg | 480 |
| | acgataatcg | agagtggcaa | tcaggacggt | tataaagggg | gtatcagcaa | tggaacgaca | 540 |
| | attaagggcg | gtgcttcacg | cgtagagggg | gggagtgcga | atggaacact | cattgatggg | 600 |
| | ggtagccaga | tagtaaaagt | tcaagggcat | gctgatggta | caacgataaa | taagtctggc | 660 |
| 50 | tctcaggacg | tagtacaagg | aagtctggca | acgaacacaa | ccataaatgg | tggctcgacag | 720 |
| | tatgttgaac | agagcacagt | agaaacaacc | accatcaaaa | atggcggtga | gcaaagagta | 780 |
| | tatgagagcc | gtgcgctgga | cacgacgatt | gaaggcggaa | ctcagtctct | gaatagtaag | 840 |
| | tcaacggcaa | aaaataactca | gatctattct | gggtgtacgc | aaattattga | taacaccagc | 900 |
| | tcctcggatg | ttattgaagt | ttattccggg | ggcgtgcttg | atgttagtg | tggtacggca | 960 |
| 55 | acaaatgtta | cccagcacga | tggtgcaatt | ttaaaaacta | acactaacgg | tacgacgggtg | 1020 |
| | agcggtagca | atagtgaagg | tgcattctcc | atccacaatc | acgtggcaga | caatgtgttg | 1080 |
| | ctggaaaacg | gtgggtcattt | agacataaac | gcatatgggt | cggcaaacaa | gacgattatt | 1140 |
| | aaagataaag | gaacaatgtc | agtttttaacc | aatgctaaag | ctgatgcgac | ccgaatagat | 1200 |
| | aatggcgggg | ttatggatgt | tgcaggaaac | gcgacaaata | ccataattaa | tgggtggcaca | 1260 |
| 60 | cagaatatta | ataattatgg | catagccaca | ggcaccaata | tcaacagcgg | aacgcaaaat | 1320 |
| | atcaaaagcg | gcgggaaagc | tgacacaaca | attatatcct | ccgggagccg | gcagggttgtt | 1380 |
| | gagaaagatg | gtacggcaat | tggcagcaat | attagcgccg | gaggctcgct | gattgtctat | 1440 |
| | accggcggta | ttgcacatgg | ggttaaccag | gagacgggca | gtgctttagt | tgccaacacg | 1500 |
| | ggtgcaggga | ctgatatcga | aggatacaac | aagctctctc | acttcactat | taccggaggg | 1560 |
| 65 | gaggctaatt | atgttgtgct | ggaaaaatacc | ggcgaactga | cggtagtgcc | taaaacctcg | 1620 |
| | gcgaaaaata | ctaccattga | tgctggcggt | aagctgattg | tccagaagga | ggctaaaaca | 1680 |
| | gatagcacca | gacttaataa | tggcggcggt | ctggagggtc | aggacgggtg | tgaggctaag | 1740 |


```

catgttgagc aacaatccgg cggcgcatta attgcttcca cgacctccgg aacacttata 1800
gaaggaacca acagttatgg tgatgctttc tacatcagga attcagaagc taaaaatgta 1860
gtgctggaaa acgctggctc attaacagtc gtcactgggt cccgggcagt tgacacgatt 1920
attaatgcca acggcaaaat ggatgtttat ggaaaagatg ttggcactgt actcaatagt 1980
5 gctggcacc ccc aaacaatata tgccagtgcc acttctgata aagcaaatat caaagggtggc 2040
aagcaaacgg tatatgggtt agccactgaa gcaaatatcg aaagtgggtga acaaatgttt 2100
gatgggtggg caacagagaa aacacacatc aatgggtggc cgcaaacctg tcagaattat 2160
ggtaaggcga tcaataccga tatcgtctct ggctacaac aaattatggc aaacgggaca 2220
gcggaagggt ccattattaa tggcgggttc cagatagtta atgagggcgg tctggctgaa 2280
10 aactcgggtg ttaatgatgg cggcacactc gatgtgcggg agaaaggcag cgcaacgggg 2340
atacagcaga gtagccaggg cgcgttgggt gcaaccacca gggcgacgcg ggtcacagga 2400
acacgcgcgg atggcgtcgc gttcagcatc gagcaggggt cggcgaaaca tatcctgctg 2460
gcaaatggcg gagtgttaac cgtggagtca gacacctctt ctgacaaaac acaggtcaat 2520
acgggcgggc gggagatcgt caaaaacaaa gccactgcga cgggcacgac-gctcaccggc 2580
15 ggtgaacaaa ttgtcgaggg tgtggcgaat gagacaacaa ttaacgacgg cggaatacaa 2640
acagtttcag ctaacggaga ggcaataaaa acaacgatca atgaaggcgg tacgctgaca 2700
gtcaacgata atggcaaagc gacagatata gtccagaaca gcggtgccgc tctccagacg 2760
agcacggcta acgggtattga aatcagcggg actcaccagt acggcacttt ttccatttcc 2820
ggcaatttag cgaccaatat gttgctggaa aatggcggta atttattggg attagcaggt 2880
20 accgaagctc gcgactccac ggttggcaag gggggggcaa tgcaaaacca gggtcaggac 2940
tccgccacaa aggttaactc tgggtggcaa tatacccttg ggcggtcaaa agatgagttt 3000
caggctctgg cccgggcaga agatctccag gttgctggcg ggacagcaat cgtctacgca 3060
ggtacgctgg cggatgcac ggtcagtggc gcgacaggaa gcctgtcgtt aatgacgcca 3120
cgggataatg ttacgccagt taaactcgaa ggggcgatcc ggattaccga tagcgcgaca 3180
25 ttaactatcg gcaatggcgt tgatacgacg cttgccgacc tgacggctgc cagccggggc 3240
agtgtctggc ttaacagcaa taattcctgt gcaggcacca gcaactgcga gtatagagta 3300
aacagtttgc tacttaacga cggtaatgtt tatttatcag cacaacagc agcgctgcc 3360
acaactaacg gtatatacaa tacgctgaca accaatgaac tttccggtag cggtaatttc 3420
tacctgcata ccaacgttgc aggtctcctg ggcgatcaac tggctcgtca caacaacgcc 3480
30 actggtaatt ttaaaatctt tgttcaggat accggcgtca gtctcagtc tgacgacgcg 3540
atgacgctgg tgaaaacagg gggaggggat gcttcgtttt cgctgggcaa tactggcggt 3600
ttcgttgatc ttgggaccta tgagtatgtc ctgaaaagcg atggcaacag caactggaac 3660
ctgaccaatg atgtcaaacc caaccggat cccaacccaa atcccaacc aaatccgaag 3720
35 ccgatccaa aaccagaccc aaaaccggat ccgaaaccag acccgactcc cgagccaacg 3780
ccgacaccgg ttccggagaa acgcatcacg ccttctaccg cagccgtact caatatggca 3840
gcaacattac cgttgggtatt tgatgctgag ctaaacagta ttccgcgagc gttgaacata 3900
atgaaagcga gtccacacaa caataatgtc tggggggcga cgtataacac ccgtaataat 3960
gtcaccaccg atgcgggggc cgggtttgag cagacgctga ccggaatgac agtggggatc 4020
gacagcccta atgatattcc tgaggggatt gcgacgctgg gcgcttttat gggttattcc 4080
40 cattcacata tcggttttga tcgcggagga catggcagtg tgggcagtta ttctctgggc 4140
ggctatgcca gttgggaaca tgaaagtggg ttctatctgg acggtgtcgt gaagctgaac 4200
cgttttgaaa gtaacgtagc cggtaaaatg agcagcgggt gagccgcaa tggcagttac 4260
cacagcaacg ggctgggcgg tcacattgaa accgggatgc gatttaccga tggtaactgg 4320
aacctgacgc cgtatgcac gttaacgggg ttcaccgctg ataaccocga atatcattta 4380
45 tccaatggca tggaatcgaa atcagtcgat acccgagta tatatcgtga actgggcgca 4440
acgctgagtt acaacatgcg tctgggggaa ggtatggaaa ttgagcgtg gctgaaggcg 4500
gctgtgcgca aagaatttgt cgatgataac cgggtgaagg tgaataatga cggtaatttc 4560
gtcaatgatt tgtcgggcag acgtggaata taccaggcag gtattaaagc ctcatcagc 4620
agtacgttaa gcgggcacat tgggggtggg tatagccatg gtgcccgtgt ggaatccccg 4680
50 tggaaacgcg tagctgggtg gaactggctg ttctga 4716

```

<212> Type : DNA

<211> Length : 4716

SequenceName : SEQ ID 386

SequenceDescription :

55

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

60 atggcagtaa agatttcagg tgtactgaaa gacggcacag gaaaaccggg agagaactgc 60
accattcaac tgaaagccag acgtaacagc gccacggtgg tgggtgaacac ggtggcctct 120
gaaaatccgg atgaagccgg tcgttacagc atggacgttg agtacggtca gtacagcgtt 180
attctgttgg tggaggggtt cccgccgtca catgccggga ccatcacctg gtatgaagat 240
tctcaaccgg gtacgctgaa tgattttctt ggtgccatga ctgaggatga tgtccgtccg 300
65 gaggcactgc gccgttttga gctgatgggt gaagaggtgg cgcgtaacgc gtccgcgggtg 360
gcacagaaca cggcagccgc gaagaagtca gccagcgatg ccagcacatc agcccgtgag 420
gcggcaaccc atgcgactga tgetgcggac tcagcacgcg cagccagcac gtcagccgga 480

```



```

caggccgcgt cgtcgggtca gtcagcgtct tccagcgcag gaacggcctc aacaaaggct 540
actgaagcat caaaaagtgc tgccgctgca gagtcctcaa aaagcgcggc ggctaccagt 600
gccgggtgcgg cgaaaacgtc agaaacgaat gcggcagtggt cacaacaatc agccgccact 660
tctgcatcca ccgcgaccac gaaagcgtca gaagctgcct cctcagccag ggatgcgtcg 720
5 gcttcaaaag aggcggcaaa atcatcagaa acgagcgcag cctcagcgc cagtagtgca 780
gcctcctcgg caacggcggc aggcaattcc gcgaaggcgg ccaaaaagtc tgagacaaac 840
gctaagtcct ctgaaacggc agcagaacag agtgccctcg cagcagcagg ctcaaaaaca 900
gcggctgcat tatctgccag tgccgctgca acaagtgcg ggccaggcctc agccagtgcc 960
accgccgccg gaaaatcggc agaaagtgcc gcatcgtctg cttcaacagc cacaacgaag 1020
10 gctggcgaag ccactgaaca ggccagcgca gcagcgagtt ctgcttccgc agcgaagaca 1080
tccgaaacga acgcgaaagc gtcggaaacc agcgcagaat cctcaaaaac ggctgccgca 1140
tcgtcagcca gttcggcggc gtcacggcca tcatctgcgt ctgcttcaaa agatgaggcg 1200
accagacaag cgtcagcagc gaagagcagc gccacgacgg catccaagaa ggcgacagag 1260
gcagctggta gtgcgacggc agcagctcag agcaaaagta cggcgggaatc tgcagcaacg 1320
15 cgcgctgaga cagcggcaaa acgggcagag gatattgcat ccgccgtggc gcttgaggat 1380
gcgagcacga cgaaaaaggg gatagtagag ctcagcagtg cgactaacag cacttccgag 1440
tactggcggc caacgccaaa agccgttaag gccgcgtatg agctggctaa cgggaaatac 1500
accgcacagg atgcaacgac agcacagaaa gggatagttc agcttagcaa cgcgaccaac 1560
agcacatctg aaatgctggc ggcaacgcca aagtcggtaa aggcagccta tgaccttgct 1620
20 aacgggaaat atactgctca ggacgctacg acagcacaac aaggaaattgt ccagctcagt 1680
agtgaacca acagcgcctc tgaaacgctt gccgcgacac cgaaagcagt gaaagcagct 1740
aatgataatg cgaatggtcg ggtaccttct gcccgtaagg tgaatggtaa ggcgctttca 1800
tcggatataa cactgacgcc gaaagatatt ggtacgctta actcaacaac aatgtcattc 1860
25 agcgggtggg ctggttggtt caaattagca acggtaacca tgcacagggc gagttctggt 1920
gtttcaatta cgttgattgg tggcgcgga ttaacgtgg ggtcacctca acaggcaggt 1980
atatctgaac ttgttttgcg tgcaggtaat ggtaatccga aggggaattac tgggtgctta 2040
tggcagcgca catcgacagg gtttacaat tttgctggg tcaatacatc tggtgatact 2100
tactgataat acgttgcaat cggaaattat gcgactgggt taaataattc atgggattat 2160
30 accagtaatg ccagcgtgac gattcatacg tcaccagcat attctgctaa taagccggaa 2220
gggttaacgg acggtacagt ttattcactc tatacgccat cagagcagtt ttatccgcct 2280
ggcgcaccaa tcccgtggcc atcagatacc gttccgtctg gctatgccct gatgcagggg 2340
cagacttttg acaaatctgc ataccgaaa cttgcagccg cttatccgtc aggcgtgatc 2400
cctgatatgc gtggctggac gattaagggc aaacctgccg gtggtcgggc cgtattgtct 2460
35 caggaacagg acggcattaa atcgcacacc cacagcgcca gcgcacagc tacggatttg 2520
gggacgaaaa ccacatcgtc gtttgattac ggcactaaat ccacgaataa caccggggcg 2580
cacacgcaca gtgtgagcgg tacagccgca agtgccggaa accatactca tagtgtcaca 2640
ggcgcatcag cagtcagcca gtggtcacaa aatgggtcag tacataaggt agtgtctgcg 2700
gccagtgtga atacaagtgc tgcaggagcg cacactcata gtgtcagcgg cacagctgca 2760
tctgcaggtg ctcacgcaca tactgtcggt attgggtgctc atacgcactc tgttgcgatt 2820
40 ggctcacatg gacacaccat caccgttaac gctgcgggta acgcggaaaa cactgtcaaa 2880
aacatcgcat ttaactacat tgtgaggctt gcataa 2916

```

<212> Type : DNA

<211> Length : 2916

SequenceName : SEQ ID 387

45 SequenceDescription :

Sequence

```

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaacgag ttattaccct gtttgctgta ctgctgatgg gctgggtcgg aaatgcctgg 60
tcattgcct gtaaaaccgc caatggtagc gctatcccta ttggcgggtg cagcgctaat 120
gtttatgtaa accttgcgcc tgccgtgaat gtggggcaaa acctggtcgt agatctttcg 180
acgcaaatct tttgccataa cgattatccg gaaaccatta cagactatgt cacactgcaa 240
55 cgaggctcgg cttatggcgg cgtgttatct aatttttccg ggaccgtaaa atatagtggc 300
agtagctatc catttccgac caccagcgaa acgcccgggg ttgtttataa ttcgagaacg 360
gataagccgt ggccgggtggc gctttatttg acgctgtgga gcagtgcggg cgggggtggc 420
attaaagctg gctcattaat tgccgtgctt attttgcgac agaccaaaaa ctataacagc 480
gatgatttcc agtttggtgt gaatatattac gccataaatg atgtggtagt gcctactggc 540
60 ggctgcgatg tttctgctcg tgatgtcacc gttactctgc cggactacct tgggtcagtg 600
ccaattcctc ttaccgttta ttgtgcgaaa agccaaaacc tgggggatta cctctccggc 660
acaaccgcag atgcgggcaa ctcgattttc accaataccg cgtcgttttc accagcgcag 720
ggcgtcggcg tacagttgac gcgcaacggc acgattatct cagcgaataa caccggtatc 780
ttaggagcag taggaacttc ggccggttaag ctgggattaa cggcaaatca cgcacgtacc 840
65 ggccgggcagg tgactgcagg gaatgtgcaa tcgattattg gcgtgacttt tgtttatcaa 900
taa 903

```

<212> Type : DNA

<211> Length : 903
SequenceName : SEQ ID 388
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| atgggctacg | ttacaggtgg | attaccaatg | aagaataacc | gtgcgtgggc | gcttatcagt | 60 |
| gggtctgatat | tggtcagcgg | aacggcccca | gctgccgata | acctgcattt | taccggtaat | 120 |
| ttgcttggtta | aatcctgtac | tcctgtaatc | aatggcaact | tacttgcaga | aattcatttc | 180 |
| cccacaattg | ctgccagcga | tttaatgcaa | cgtggtcagt | cagatcgcgt | accgttagtt | 240 |
| tttcagttga | aagattgcaa | aagcaccacg | gcgtttaatg | tcaaggtgac | cttgatggga | 300 |
| acagaagata | ccgacttacc | aggatttctg | tcgattgatt | cgtcatcttc | tgcaacgggt | 360 |
| ggtgggattg | gcattgaaac | tgccggaggg | gcggctgtac | ctattaacag | taccacaggt | 420 |
| gcctcatttc | cattaaatca | gggaaataac | agtgtcaatt | ttaatgcctg | gttacagacc | 480 |
| gtaaatggac | gaaatgttac | atcgggtgat | ttcaccgcca | caatgacggt | aacttttgag | 540 |
| tattttttaa | | | | | | 549 |

<212> Type : DNA

20 <211> Length : 549

SequenceName : SEQ ID 389

SequenceDescription :

25 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

| | | | | | | |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| atgaaacgac | atctgaacac | cagctacagg | ctggtatgga | atcacattac | gggcaccctg | 60 |
| gtgggtggcct | ccgaactggc | ccgctcacgg | ggaaaacgcg | ccggtgtggc | ggttgcgctg | 120 |
| tctcttgctg | ctgtcacatc | agtcgccgga | ctggctgctg | acaaggttgt | acaggcggga | 180 |
| gaaaccgtga | acgatggaac | actgacaaat | catgacaacc | agattgtctt | cggtacggcc | 240 |
| aacggaatga | ccatcagtag | cgggctggaa | ctggggcccg | acagtgaaga | aaacaccggt | 300 |
| gggcaatgga | tacagaatgg | cgggatagcc | ggaaacacca | ctgtcaccac | aaatggtcgt | 360 |
| caggctcgtg | tggagggggg | aacagccagt | gatacggtta | ttcgtgacgg | cgggggacag | 420 |
| agcctgaacg | gactggcggg | gaacaccaca | ctgaataaca | gaggcgagca | gtgggtgcat | 480 |
| gagggcgggg | ttgccaccgg | tacaattatc | aaccgcgacg | gttaccagag | cgttaaaagt | 540 |
| ggcgggctgg | caacaggaac | catcatcaac | accggcgacg | aaggcggccc | tgattctgac | 600 |
| aactcgtata | cgggtcagaa | ggtccaggga | acagcagaat | ccaccaccat | caacaaaaat | 660 |
| ggacggcaga | ttatcttatt | ttccgggcta | gcccgtagca | ctctcattta | cgcagggtgg | 720 |
| gaccagtcgg | tacacggaag | ggccctgaat | accacactga | atggcgggta | ccaatatgtg | 780 |
| cacagggacg | gacttgcgct | gaacacggta | attaacgagg | ggggctggca | ggttggttaag | 840 |
| gcaggtggcg | ctgccggtaa | caccaccata | aatcagaacg | gtgaactgag | ggtacatgcc | 900 |
| ggcggggaag | ccactgcagt | caccacgaac | acgggcggtg | cactgggttac | cagtactgct | 960 |
| gcaactgtca | tcggcacaaa | ccgtctgggg | aatttcacgg | tggaaaacgg | taaggctgac | 1020 |
| ggtgtgtgtc | tggaatccgg | cggctcgtctg | gatgtactgg | agagccattc | agcacagaat | 1080 |
| accctagtgg | atgacggcgg | taccctggca | gtgtctgccg | gcggtaaggc | gacaagtgtc | 1140 |
| accataacat | ccggtgggtg | cctgattgca | gacagtgggt | ccactgttga | ggggaccaat | 1200 |
| gccagcggta | agttcagtat | tgatggcaca | tcgggtcagg | ccagcggcct | gctgctggaa | 1260 |
| aatggcggca | gctttacggt | taatgccggg | ggacaggctg | gcaacaccac | tgctcgacat | 1320 |
| cgtggaacac | tgacgctggc | tgccggggga | agtctgagtg | gcagaacaca | gctcagtaaa | 1380 |
| ggcgccagta | tggtactgaa | tggtgatgtg | gtcagtaccg | gcgatattgt | taacgcaggg | 1440 |
| gagattcgct | ttgataatca | gacgacaccg | aatgccgcgc | tgagccgtgc | tgttgcaaaa | 1500 |
| agtaactccc | cggtaacggt | ccataaactg | accaccacga | acctcaccgg | ccagggcggc | 1560 |
| accatcaata | tgcgtgttcg | ccttgatggc | agcaatgcct | ctgaccagct | ggtgattaat | 1620 |
| ggtggtcagg | caaccggcaa | aacctggcct | gcgtttacaa | atgtcggaaa | cagcaacctc | 1680 |
| ggggtggcaa | ccaccggaca | gggtatccgg | gttgtggatg | cacagaatgg | cgccaccaca | 1740 |
| gaagaagggt | cgtttgccct | gagtcgcccc | cttcaggccg | gcgcctttaa | ctacaccctg | 1800 |
| aaccgtgaca | gcgatgaaga | ctgggtacctg | cgcagtgaac | atgcttatcg | tgctgaagtc | 1860 |
| cccctgtata | catccatggt | gacacaggca | atggactatg | accggattct | ggcaggctcc | 1920 |
| cgcagccatc | agaccgggtg | aaacgggtgaa | aataacagcg | tccgtctcag | cattcagggc | 1980 |
| ggtcatctcg | gtcacgataa | caacggcggg | attgcccgtg | gagccacgcc | ggaaagcagc | 2040 |
| ggcagctatg | gcttcgctccg | tctggagggg | gacctgctca | gaacagagggt | tgccggtatg | 2100 |
| tctctgacga | caggggtgta | tggtgctgca | ggccattctt | ccgttgatgt | taaggatgat | 2160 |
| gacggttccc | gcgcgggca | ggcccgggat | gatgcgggca | gtctgggagg | atacctgaat | 2220 |
| ctggtacaca | catcctccgg | cctgtgggct | gacattgtgg | cccagggaac | ccgtcacagc | 2280 |
| atgaaagcgt | catcggacaa | taacgacttc | cgcgcgccgg | gctgggggctg | gctgggctca | 2340 |
| ctggaaaccg | gtctgccctt | cagtatcact | gacaatctga | tgctggagcc | acaactgcag | 2400 |

tacacctggc agggactctc cctggatgac ggccaggata acgccgggta tgtgaagttc 2460
gggcatggca gtgcacaaca tgtgcgtgcc gggtttccgtc tgggcagcca caacgatatg 2520
acctttgggtg aaggcacctc atcccgtgac accctgcgcg acagtgcaaa acaacagtgtg 2580
agtgaactgc cgggtgaactg gtgggtacag ccttctgtta tccgcacctt cagctcccgg 2640
5 ggtgacatga gcatggggac agccgcagcc ggcagtaaca tgacgttctc accgtcccgg 2700
aatggcacgt cactggacct gcaggccgga ctggaagccc gtatccggga aaatatcacc 2760
ctgggcgttc aggccgggta tgccacagc gtcagcggca gcagcgtga aggcataaac 2820
ggtcaggcta cgctgaatat gactttctga 2850
<212> Type : DNA
10 <211> Length : 2850
SequenceName : SEQ ID 390
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaaaat ggcattatat attttgcata atttcttttc atttaggggt accgtgcggg 60
tatgcggcaa atgatggcac gtgtgcaaca agaggcggca cacatacatt aagccttaat 120
20 tttcctctga caacggtcag tgcagcaaac aatgtgcctg gaaatacatt aatagatatt 180
gctaatagcaa catcttctga aaattatagc gttctgtgta actgtgattc aaaacatagc 240
aatggcgctt atcacgaaat atattatacc gcagaccctg ctcccggat ggtttatagc 300
accaccgcaa gtggtcttgc tttttactat cttaacgaat atgtcgatgt gggaacaaaa 360
atatctgtgc taaatgcggg gtatacggca gttccttttg aacatgtttc caaccaggca 420
25 actacaacag atcacacttg tcagggaaac aaaactacag cggttggcgt gagcctgaaa 480
actggagcag atgcgaagat ttcatttcgt attaaacgtt caataaatgg aacggtagta 540
atacctatca ccgataattgc attgctgtat gccaacatat ccagcaccac gaccctgggt 600
gaggcgattg caaaagttcg aatttcaggc agtttgaccg caccacagtc ttgtcagata 660
aatgcaggac aggtgattta ttttgatttt gatactattc ctgcgtccga attttcatct 720
30 accgccgggc aagccattac ttcacgaaaa atcactaaaa cagttagtat tgagtgtacg 780
gggatggggg atgagcgtac gcagaaagtc gatgcttctt ttacggggac gaaccgaagc 840
agtgcagata cgatgggtggc gacagacaat gctgatgtcg ggatcaaaat ttacaataaa 900
tcgaatgctg aagtttagcgt caacaacggc aagttaccgc cagacatggg caacacgacc 960
atttttggtc gtaaaaatgg ttcggtaact ttttcggcag cacctgccag ctttaccggg 1020
35 gcccggcctc agcccggcgt ttttaacgct accgcgacct taaccattga atttgtaaac 1080
taa 1083
<212> Type : DNA
<211> Length : 1083
SequenceName : SEQ ID 391
40 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
45 <400> PreSequenceString :
atgtcacgtt ataaaacagg tcataaacia ccacgatttc gttattcagt tctggcccgc 60
tgctggcgt gggcaaatat ctctgttcag gttctttttc cactcgtgt caccctttacc 120
ccagtaattg cggcacgtgc gcagcatgcg gttcagccac ggttgagcat gggaataact 180
acggtaactg ctgataataa cgtggagaaa aatgtcgcgt cgtttgccgc aaatgccggg 240
50 acatttttaa gcagtcagcc agatagcgat gcgacacgta attttattac cggaaatggc 300
accgctaaag ctaaccagga aatacaggag tggctcggga aatatggtac agcgcgcgtc 360
aaactgaatg tcgataaaga tttctcgtc aaggattctt cgctggaaat gctttatccg 420
atztatgata cgccgacaaa tatgttggtc actcaggggg caatacatcg tacagacgat 480
cgtactcagt caaatattgg ttttggtcgt cgtcattttt caggaaatga ctggatggcg 540
55 ggggtgaaca cttttatcga ccattgattt tcccgtagtc ataccgcgt tgggtgttgg 600
gcggaatact ggcgcgatta tctgaaactg agcgcgaatg gttatatccg ggcttctggc 660
tggaaaaaat cgccggatat tgaggattat caggaaacgc cggcgaatgg ttgggatata 720
cgcgacagag gctattttacc tgccctggcg cagcttggcg caagcctgat gtatgaacag 780
tattatggcg atgaagtcgg gctgtttggg aaagataagc gccagaaaga ccgcgatgct 840
60 atttctgccg aggtgacctt tacgccagtg cctcttctga cactgagcgc cgggcataag 900
cagggcaaga gcggtgagaa tgacactcgc tttggcctgg aagttaacta ccgaattggc 960
gaaccttttg cgaaacaact cgatacggat agcattcgcg agcgtcgggt actggcaggc 1020
agccgctatg acctggttga gcgtaataac aacatcgttc ttgagtaccg caaatctgaa 1080
gtgatccgta ttgctctgcc tgagcgtatt gaaggtaagg gcggtcagac actttccctg 1140
65 gggcttgttg tcagcaaagc aactcacgga ctgaaaaatg tgcagtggga agcgcctca 1200
ttactggctg aagggtggca aattaccggg cagggtagtc agtggcaagt aacgctcccg 1260
gcttatcgtc caggcaaaaga caattattat gcgatttcag cagttgcta cgataaaaa 1320


```

5  ggcaatacct caaaacgcgt gcagacagag gtggtcatta ccggagctgg tatgagcgcc 1380
   gatcgcacgg cgttaacgct tgacggtcag agccgtattc aaatgcttgc taacggtaat 1440
   gagcaaaaac cgctgggtgct gtctctgccc gacgcccagg gccagccagt cacggggcatg 1500
   aaagatcaga tcaagactga actaactttc aaaccggctg gaaatattgt gactcgttcc 1560
   ctgaaggcca ctaaatacaca ggcaaagcca aactgggtg agttcaccga aactgaagca 1620
   ggggtgtatc agtctgtctt tactaccgga acgcagtcag gtgaggcaac gattactgtt 1680
   agcgttgatg gcatgagcaa aaccgtcact gcagaactgc gggccacgat gatggatgtg 1740
   gcaaactcca ccctgagcgc taacgagccg tcaggtgacg tggttgctga tggtcagcaa 1800
   gcctatacgt tgacgttgac tgccgtggac tccgagggtg atccgggtgac gggagaagcc 1860
10  agccgcttgc gatttgttcc gcaagacact aatgggtgtaa ccgttgggtg catttcggaa 1920
   ataaaaccag gcgtttacag cgcgcgggtt tcttcgaccc gtgcccggaaa cgttgttgtg 1980
   cgtgctttca gcgagcagta tcagctgggc acattacaac aaacgctgaa gtttgttgcc 2040
   gggccgcttg atgcagcaca ttcttccatc accctgaatc ctgataaacc ggtgggtggg 2100
   gggacagtta cggcaatctg gacggtaaaa gatgcctatg acaaccctgt gaccagcctc 2160
15  acgcccgaag cgcgctcatt agcgggtgcc gctgctgaag gttctacggc atccgggtgg 2220
   acaataaatg gtgatgggac gtggactgcg cagattactc tcggctctac ggcgggtgaa 2280
   ttagaagtta tgccgaagct aaatggacag aatgcggcag caaatgcggc aaaagtaacc 2340
   gtgggtggctg atgcgttatc ttcaaaccag tcgaaagtct ctgtcgcaga agatcacgta 2400
   aaagccggcg aaagcacaac cgtgacgctg gtggcgaaag atgcgcatgg caacgctatc 2460
20  agtgggtctt cgttgtcggc aagtttgacg gggaccgcct ctgaaggggc gaccgttcc 2520
   agttggaccg aaaaaggtaa cggttcctat gttgctacgt tgactacagg tggaaagacg 2580
   ggcgagcttc gcgtcatgcc tctcttcaac ggccagccag cagccaccga agccgcgcag 2640
   ttgacgggtc tcgcccggaga gatgtcatca gcgaactcta cgcttgttgc ggacaataag 2700
   gctccgaccg tcaaaacgac gacggaactc accttcaccg tgaaggatgc gtacgggaac 2760
25  ccggtcaccc ggctgaagcc agatgcacca gtgttttagc gtgcccggcag cacggggagt 2820
   gagcgtcctt cagcaggaaa ctggacagag aaaggtaatg gggctctact gtcgacctta 2880
   acgctgggat ctgccgcggg tcagttgtct gtgatgccgc gactgaacgg ccaaaatgcc 2940
   gttgctcagc cactgggtgt gaacgttgca ggtgacgcac ctaaggctga gattcgtgat 3000
   atgacagtga aggttaataa ccaactggct aatggacagt ctgctaacca gataaccctg 3060
30  accgttgttg acacctatgg taaccgcttg caggggcagg aagttacgct gactttaccg 3120
   cagggtgtga ccagcaagac ggggaataca gtaacaacta atgcggcagg taaagcggac 3180
   attgagctta tgtcaacggg tgccgggagaa cacaatatct ccgcttcggg gaatggtgct 3240
   cagaagacgg tcacgggtgaa attcaacgcg gatgccagca ccggtcaggc aaacctgcag 3300
   gtagacgccg ctgctcaaaa agtggcaaac ggcaaagatg cctttacgct gacggcgaac 3360
35  gttgaggata aaaatggtaa ccctgttcca gggagcctgg tgacctttaa tctgccccgg 3420
   ggtgtcaagc cgcttacagg cgataatgtc tgggtgaaag ccaacgatga ggggaaagca 3480
   gagttgcagg tggtttcagt gactgccgga acgtatgaga tcacggcatc ggcagggaat 3540
   agccagcctt cgaatacgcg gactataacg tttgtagccg ataaggctac cgcaaccgtc 3600
   tccggtattg aggtgattgg caactatgca ctggcggacg gcaatgccaa acagacgtat 3660
40  aaagttacgg tgactgatgc caataacaac ctgttgaaag atagcgaagt gacgctgact 3720
   gccagcccgg caaatttagt tctgactccc aatgggacgg cgaaaactaa tgagcaagga 3780
   caggctatct tcaccgccac gaccactgtc gcagcgaaat atacactcac ggcgaaagtg 3840
   agtcaggccg acggtcagga atcgacgaaa actgccgaat ctaaatcgtc cgcgatgat 3900
   acaaatgcag tactaccgcg atcatctgat gtgacttctc tgggtggcggg tgggatatcg 3960
45  actgcgaagc tggaggtgac actgatgtcg gcaaataacc ccgttggggg gaatatgtgg 4020
   gtcgacatta agacgccaga aggggtgacg gagaaggatt atcagttcct gccgtcgaaa 4080
   aatgaccatt tcgtgagcgg aaaaatcacg cgtacattta gtaccagcaa gcctgggtgtc 4140
   tatacgttca catttaacgc cctgacgtat ggcgggtacg aaatgaagcc agtgacgggtg 4200
   accattaccg cgggtgatgc cgatacggca aagggcgagg aggcgatgaa ctaa 4254

```

<212> Type : DNA

<211> Length : 4254

SequenceName : SEQ ID 392

SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

60  atggcgcggtg gttgggcgtc ttcagaagcc tcaggcgcca tgactgattg gttaaataac 60
   tttggtaactg cgagaatctc tctgggtgtg gatgaagatt ttagcctgaa aaattcgcaa 120
   ttcgacttcc tgcattccgtg gtatgacaca cctgattatc tgctcttcag ccagcatacc 180
   cttcacccgaa cagacgatcg taaccagatc aacaccgggt tgggctggcg tcatttcacc 240
   tccagctgga tgtcaggcat caaccttttt tttgaccacg acctgagccg ctatcactcc 300
65  cgcgccagggc ttggcgcgaga atactggcgt gatttatctga agttgagcag caacgcttat 360
   atcggcctga ccggctggcg tagcgcacca gaattggata acgacttcga agcccggccg 420
   gccaacggct gggatttacg cgcggaaggg tggttacctg cctggccaca actgggggga 480

```

| | | | | | | | |
|----|-------------|-------------|-------------|-------------|-------------|-------------|------|
| | aaactgggtct | atgaacaata | ctatggcgat | gaagtggcgc | tgtttgacaa | gaatgatcgt | 540 |
| | caaagtaacc | cccatgctat | taccgcaggc | ctcaactata | cccccttccc | gcttctgact | 600 |
| | ctcagtgccg | aacagcgtca | ggggaagcaa | ggtgaaaatg | acacacgttt | tgccgttgat | 660 |
| | ctgacctggc | aaccagcag | ttcaatgcag | aaacagctta | atccggacga | agtggccgga | 720 |
| 5 | cggcgcagtc | tggccggtag | tcgttatgac | ctgattgatc | gcaacaacaa | catcgttctg | 780 |
| | gaataccgca | agaaagagct | gattcgcttg | agtctgctgg | atccggtgaa | aggggaagtct | 840 |
| | ggagaaataa | aaccgctggg | ttcctcgcta | cagaccaaata | atgcccttaa | aggctataac | 900 |
| | atcgaagccg | ctgcgctgga | agctgccgga | ggtaaagtca | gcacgtctgg | aaaagatatc | 960 |
| | acggtcacgc | tgccagggtta | ccgcttcact | aacaccccag | aaaccgataa | tacatggctcg | 1020 |
| 10 | atagacgtta | ccgccgagga | tgtaaaagggt | aacctgtcac | ggcatgaaca | aagcatggta | 1080 |
| | gttattcagg | ctccgacatt | aagccagaaa | gattctctgt | tatccgtcaa | tccgctaacc | 1140 |
| | gtggctgcag | ataaaaaatc | gacgaccaca | ttgaccgtta | ctgcgcacga | ttccgacgga | 1200 |
| | actccgggtgc | cggggcctggc | gctgcaaacc | cgcagtgaag | gcgttcagga | tatcacctcg | 1260 |
| | tctgactgga | cagataacgg | tgatggtagt | tacacacaga | tactgacgcg | cggaaacgaca | 1320 |
| 15 | tcaggttcag | taacactgac | gccgcaaatt | aacgggtgaga | gtgcggtaaa | agaatccatc | 1380 |
| | gtcgttaata | tcgtccctgt | tgtctcatcc | cgcgaccatt | catcaataac | aattgataac | 1440 |
| | gtatcgtatt | atgccggaga | cgacatcaag | gttaggggtg | aactgaaaga | cgatagcaat | 1500 |
| | caaccgggttg | catatcaaaa | agaggaattg | gtaaaagccg | ttactgtcga | aaacagcaaa | 1560 |
| | cctggcgcca | cgattgtctg | gcacgaagag | cagccggggg | tttatgccgc | gaattatccg | 1620 |
| 20 | gcctataagc | aagggactgc | actaagggca | caacttagcc | ttcacaactg | gaatgctcca | 1680 |
| | ctgcaatcgc | atattttataa | cattgaggca | aaccagaata | aggctcgcgt | tgccacatta | 1740 |
| | tcagcgacaa | ataatgacgt | ttacgccgat | aaaaagacat | ttaataccct | cacgatcaac | 1800 |
| | gtcactgatg | agagtgataa | ttccctgaca | aatcatcagg | tcacctttta | gaatgaaaaa | 1860 |
| | ggaagcgcgg | agtttgtcga | accgccgcag | caaaatacgg | atgcatatgg | tggttgccaca | 1920 |
| 25 | ataaacatgg | taagtcagggt | tgcggaagaa | aatacagatta | gcgccacgct | gccaaatggg | 1980 |
| | ttttcacaa | ggataattgc | gaaattcgtt | agcgattcga | gtacgccaaa | attcaaacaa | 2040 |
| | ctgggttgccg | atccagatac | cattattgct | ggcaacagcc | agggcagtac | tctgaccgcc | 2100 |
| | atcatcacag | actttcataa | caaccggtta | aaagatatga | aagtgaattt | tgtggcacct | 2160 |
| | ggtggctcgc | aactggacaa | cacgaccgcc | acaacagacc | agtcgggtat | tgtgcgggtg | 2220 |
| 30 | cacctgacca | gttcaaaagc | tggtagctat | tcctgcgatg | cctcgttga | ggtggataaa | 2280 |
| | aatattcacc | agtcgggtcac | gatcaccttg | gtcccaaaca | gggaacaatc | ggtaatgacc | 2340 |
| | ttgaatgccg | ggtcggggcag | tgcgatcgc | aacaatacaa | atatcgttac | cctgactgcc | 2400 |
| | agtgtgaaag | atgtttatgg | acaccggttg | ccggatgagg | atgtgaaatt | taccttgcca | 2460 |
| | gcctccatga | ccgggaactt | cacgctaagt | agtgaaccgc | ccgcaccga | tgcaaacggg | 2520 |
| 35 | gatgcgctgg | tcacattgctg | aggcacaata | gcgggtgagt | ttacagttac | ggcgacgctg | 2580 |
| | accagaaata | ataccggtgc | ttatcagcaa | gtcactttta | ttggggatag | aaacagtgctg | 2640 |
| | cagctccagc | cgctgactgc | ctcattaaat | ttcattgttg | cgggtaacag | tacggggagt | 2700 |
| | accctgacgg | caacgatcct | ggacgcttac | caaaatccgc | ttaaagacca | gttgggtcact | 2760 |
| | ttccagagta | acgatgtcac | tctaagcgaa | acagaagtca | ccaccaatac | gctgggtcag | 2820 |
| 40 | gcgacggtaa | caatgaccag | caatattgcc | ggacaacata | acgtcgtggg | gagccggaaa | 2880 |
| | gcgcaagctt | ccgataataa | aacgtttagt | ttatcagtgc | taccggatga | aagttcggcg | 2940 |
| | aaggtaataa | gtataaccgg | agccgaaaaa | acgataacgg | tgggcgaaaa | catcacgcta | 3000 |
| | cggataactcg | tccaggacgc | gtttaacaat | gtaatcgccg | gtcaacgcgt | cagattaaat | 3060 |
| | gcgcagccaa | caactaacat | tacgataggc | gatacggctt | acaccgataa | taacgggttat | 3120 |
| 45 | gcgtacgtta | accttctcag | cacccaacct | gggggtttatc | aggtgacggc | aacgctggac | 3180 |
| | aataacagta | gtagtaagggt | tgacgtgaat | gtggcaaatg | gcaaaactcga | gttaacatca | 3240 |
| | tcgaaaccag | aaactacggg | ccataatagt | gaggggtatta | cgctgaccgc | aacggcgaga | 3300 |
| | aatgcgcggg | gtgaattgat | gccagggcaa | attatcacct | ttagcgtaac | gcctgaagggt | 3360 |
| | gcaacgctaa | gcaatacagg | ggaagtccct | actgaccagt | caggtcaggc | caaagtgcag | 3420 |
| 50 | ctgaccagtg | acaaagtga | tgtctatacc | gttacggcca | taatgggcaa | agatgttccc | 3480 |
| | gttcagagcc | aggtaacggg | tgcgggttaag | gcagatgcta | aaacggcaca | tgttgtgagc | 3540 |
| | gtcgtggctt | ctcctgacac | catcacccgc | gacggcatcg | atagcagcac | catcacttca | 3600 |
| | cgagtagaag | atgattacgg | attcccggtt | gaagggtgctg | atattagtca | tggcttagac | 3660 |
| | accaaaggca | gcccggtagt | taatattcca | actacgcgta | ccgatcagtc | cgggcaagtc | 3720 |
| 55 | acggcgacaa | taaccagtac | attggcagaa | accttaacag | tcaatgtgca | agttcctggc | 3780 |
| | acagccaacc | aatccgcaac | cattacattg | gttgccggca | cggccgatga | aagtaagtca | 3840 |
| | atthttgaaat | ccgatgttga | cactctgaag | gctgactacc | agcagagcgc | aaaacttacg | 3900 |
| | ctaacattgc | aagacaagta | cggtaaacccg | atagtgaagt | ctgatcatct | ggaatttgtc | 3960 |
| | cagtcaggcc | ccttcgtgaa | ctttctcaag | ttgagcgata | ttgattacag | ccaaagaaat | 4020 |
| 60 | tatggcgagt | acaccgtgac | tgtcactggc | ggaaaagagg | gaacagcgac | actcattccc | 4080 |
| | atgctgaacg | gggttcacat | ggcaaaactta | agcatatcgc | tgaatctcat | ccaatcgata | 4140 |
| | aaagaaatgt | ccgggtcatgt | cactgcaaac | aaccatacct | tctccacggc | taaattcccg | 4200 |
| | agcgaaggct | ttgcaggagc | gtattacaca | ctcaacaatg | ataactttga | agcgggtaaa | 4260 |
| | accgttgatg | attatatgtt | ttcaagttca | caggggttggg | tgtctgtcga | tgcttcgggt | 4320 |
| 65 | aaagtthctt | tcgcaaatat | cggcgatcaa | acgtcagtc | caataagcgc | tgthtcccga | 4380 |
| | caaggaggta | caacctacca | gaccttaatt | aagctgaaag | gctgggtggg | gaataatgga | 4440 |
| | aatcatacca | atatctggct | agctgccaat | gcgctctgtc | atgctaaaaa | tgatggatat | 4500 |

aatcttcctg gcatcacaca tttgacgtct ggcgaaaaca aacgcacgca gggatcactg 4560
tatgggtgaat gggggaacgt tggagcggtt tccagtaatt cgcaatttac accgggagct 4620
tactggacaa gtgaatctga tgattacagt cggcactact atgtgcagat gctaaccggt 4680
atgaccggaa gcgacgctga ttccagcccc caactgaccg cctgccgtaa atcactttaa 4740

5
<212> Type : DNA
<211> Length : 4740
SequenceName : SEQ ID 393
SequenceDescription :

10
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

15 atgattactc atgggttgta taccgggacc cggcacaagc ataagctaaa aaaaacattg 60
attatgctta gtgctgggtt aggattgttt ttttatgtta atcagaattc atttgcaaatt 120
gggtgaaaatt attttaaaatt ggggttcggat tcaaaactgt taactcatga tagctatcag 180
aatcgccttt tttatacgtt gaaaactggt gaaactgttg ccatctttc taaatcgcaa 240
gatattaatt tatcgacgat ttgggtcggtg aataagcatt tatacagttc tgaaagcgaa 300
20 atgatgaagg ccgcgcctgg tcagcagatc attttgccac tcaaaaaact tccctttgaa 360
tacagtgcac taccactttt aggttcggca cctcttggtg ctgcaggtgg tgttgctggt 420
cacacgaata aactgactaa aatgtccccg gacgtgacca aaagcaacat gaccgatgac 480
aaggcattaa attatgcggc acaacaggcg gcgagtctcg gtagccagct tcagtgcgca 540
tctctgaacg gcgattacgc gaaagatacc gctcttggtg tcgctggtaa ccaggcttcg 600
25 tcacagttgc aggcctgggtt acaacattat ggaacggcag aggttaattc gcagagtggg 660
aataactttg acggtagttc actggacttc ttattacgtt tctatgattc cgaaaaaatg 720
ctggcatttg gtcaggctcg agcgcgttac attgactccc gctttacggc aaatttaggt 780
gcgggctcagc gttttttcct tccctgcaaac atgttgggct ataactgctt cattgatcag 840
gatttttctg gtgataatac ccgttttaggt attgggtggcg aatactggcg agactatttc 900
30 aaaagtagcg ttaacggcta tttccgcagt agcggctggc atgagtcata caataagaaa 960
gactatgatg agcgcgccagc aaatggcttc gatatccgtt ttaatggcta tctaccgtca 1020
tatccggcat taggcgccaa gctgatatat gagcagattt atgggtgataa tgttgctttg 1080
tttaattctg ataagctgca gtcgaatcct ggtgcggcga ccgttggtgt aaactatact 1140
ccgattcctc tgggtgacgat ggggatcgat taccgtcatg gtacgggtaa tgaaaatgat 1200
35 ctcttttact caatgcagtt ccgttatcag tttgataaat cgtggtctca gcaaattgaa 1260
ccacagtatg ttaacgagtt aagaacatta tcaggcagcc gttacgatct ggttcagcgt 1320
aataacaata ttattctgga gtacaagaag caggatattc tttctctgaa tattccgcat 1380
gatattaatg gtactgaaca cagtacgcag aagattcagt tgatcgttaa gagcaaatac 1440
ggtctggatc gtatcgtctg ggatgatagt gcattacgca gtcagggcgg tcagattcag 1500
40 catagcggaa gccaaagcgc acaagactac caggctatct tgcctgctta tgtgcaagggt 1560
ggcagcaata tttataaagt gacggctcgc gcctatgacc gtaatggcaa tagctctaac 1620
aatgtacagc ttactattac cgttctgtcg aatgggtcaag ttgtcgacca ggttggggta 1680
acggacttta cggcggataa gacttcggct aaagcggata acgccgatac cattacttat 1740
accgcgacgg tgaaaaagaa tggggtagct caggctaatt tccctgtttc atttaatat 1800
45 gtttcaggaa ctgcaactct tggggcaaat agtgccaaaa cggatgctaa cggtaaggca 1860
accgtaacgt tgaagtcgag tacgccagga caggctcgtg tgtctgctaa aaccgcggag 1920
atgacttcag cacttaattg cagtgcggtt atattttttg atcaaacc aa gccagcatt 1980
actgagatta aggtgataa gacaactgca gtagcaaatg gtaaggatgc tattaaatat 2040
actgtaaaag ttatgaaaaa cggtcagcca gtttaataatc aatccgttac attctcaaca 2100
50 aactttggga tgttcaacgg taagtctcaa acgcaagcaa ccacgggaaa tgatggctcg 2160
gcgacgataa cactaacttc cagttccgcc ggtaaagcga ctgttagtgc gacagtcagt 2220
gatggggctg aggttaaagc gactgaggtc actttttttg atgaactgaa aattgacaac 2280
aagggtgata ttattggtaa caatgtcaga ggcgagttgc ctaatatatt gctgcaatat 2340
ggtcagttta aactgaaagc aagcgggtgg gatggtacat attcatggta ttcagaaaat 2400
55 accagtatcg cactgtcga tgcacaggg aaagtcactt tgaatggtaa aggcagtgct 2460
gtaattaaag ccacatctgg tgataagcaa acagtaagtt acactataaa agcaccgtcg 2520
tatatgataa aagtggataa gcaagcctat tatgctgatg ctatgtccat ttgcaaaaat 2580
ttattaccat ccacacagac ggtattgtca gatatttatg actcatgggg ggctgcaaat 2640
aaatatagcc attatagttc tatgaactca ataactgctt ggattaaaca gacatctagt 2700
60 gagcagcgtt ctggagtatc aagcacttat aacctaataa caaaaaacc tcttcctggg 2760
gttaatgtta atactccaaa tgtctatgcg gtttgtgtag aataa 2805

<212> Type : DNA
<211> Length : 2805
SequenceName : SEQ ID 394
SequenceDescription :

65
Sequence


```
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
5 atgttagtac ttagcgaaag cttcaagaat aaattgcttc ccatgaatgg gtatatgaaa 60
  ggcggcagcg actccggatc taaagcccag gcacgcgcaa ctgaaaaggg catcgaaactg 120
    cagcgtgaaa tgtggcagac gaacatgcaa aaccttgcac cgttcacgcc actcgctcag 180
      cagtacgtat cacagttgca gaatcttctc tctcttcagg ggcaagggtca ggcgcttaac 240
        cagtattaca actctcagca gtataaagac cttgcagggc aggcgcgcta tcagagtctg 300
          gcagcagcag aggcaacggg tggattaggc tctacagcaa caggaaacca gttagcagca 360
            atcgcaccta cactcgggtca aaactggctg tcagggtcaga tgaacaacta caacaatctg 420
              gcaaatatcg gccttggtgc tcttacaggt caggcaaacg ccggacagaa ctacgctaac 480
                aatgtcagcc aattgtatca acagcaggcg gcagcatcgg cagcaaatgc gaataagcct 540
                  tcaggcctac agagtcttgc tacaggtgcc attgggtggg ccgcatcagg tgcaatgatt 600
                    ggtagtgcag ttctgtttat tgggactggg attgggtgct ttgctggcgg tgttatcggt 660
                      ggtcttggat cattgtttta a 681
<212> Type : DNA
<211> Length : 681
      SequenceName : SEQ ID 395
      SequenceDescription :
20
Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 atgaaaaaaa tattatcagg gttgattctg ctgctttgct gtccttatgg tttcgccgct 60
    aatgggtgatg ggcgaacgca catgtcaaat ttatcatttg gtccgctgac ggtggcagcg 120
      gcgaataacc actccggata caatatcttc gaggcactga gcaacacgac tggaacatac 180
        ccggtgcgct gtcactgtga tgacacgcat ggcggaccgg gccaacaaac agcatttttt 240
          cctatcttct acacgggaga tgccgcaccg gggcttgtgc ttgagcgcac tcttaatggg 300
            tttaaattact atgctctgaa tgattattta tcggtcggcg tgacgatttt tattattaat 360
              aaccaatatg cggccattcc ttttgaacac ttatccaacc aatccacctc accgcaacat 420
                acctgtggag caggtaataa tgggagcact gttaaactctg attcagggcg ctccggcaaaa 480
                  ttatctttct atgttcggca ttctattact ggcacgggtga caataccac aacggaagtc 540
                    gcctgggttg acgcgggcat gtccgatcat tttcccaaaa cgacccccgt ttctaaagtg 600
                      acaattcgcg gacaactaac ggcgccgcag aactgtgagt taacgccaaa tcagagcatc 660
                        gatgtcgatt ttcaaaaaat taatagcgct gagttctcct caacggcggg ttcaattatt 720
                          gcagaaagaa agattaaaac cgaagtcaca gtatcctgta ccgggatgga agacgtaagg 780
                            tccacggagg tggtagtgct gtcgatgatt gcggcaaaca gaagtgccga tgccaccatg 840
                              atcgtgacga gtaatccgga tgtgggaatt aagatttttg ataagaacga ccgtccagtg 900
                                aatgtggatg ggggcaactt acccgctgat atgggtgcta ttagtcgatt aggaaaaacc 960
                                  gatggtagcg taacgtttta ttcagcgccc gccagtctga cgggcgcaaa accagcgcc 1020
                                    gataatggat ttaccgctac agccacgctg gttattgaat ttactaacta a 1071
<212> Type : DNA
45 <211> Length : 1071
      SequenceName : SEQ ID 396
      SequenceDescription :
50
Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 atgaataaaa tatatcggct aaagtggaaac aggtcccgtg actggtggag cgtctgctcg 60
    gagctgggga gcagagtaaa aggaaaaaag tcccgggctg ttttaattag cgcgataagt 120
      ttatattcat ctctggtatt cgccgatgat gtcacgttaa accaggataa aactattgat 180
        tttggcaaag agaaccagag catcgattac cgtattacgg tgacagacaa tgccaatctg 240
          gtaatcaatg cgacagatac ttcccgtccg cgtctgactc tcgcttctgg tggtaggttg 300
            gatattaccg gaggaagggt aacaatcaat ggcccactta actttttgct gaaaggtagc 360
              gggttcctga atgtctccaa tgctggcagc gagttatatg ctgatgattt gtatgaatca 420
                aactcaggca tgagacacga tcgcggctat tttaatgtct ccaacggcgg caaatccat 480
                  gttaagggca ccagccgtct gacctatttg cagggaatg tcagtgggtg aggtagccag 540
                    gtaaattccg aaaccttctt tatgggcgtt tacggcagtt acggtggtga tcagtacctg 600
                      tcagttaata acggcgggtg agttaatgcc aggaagcaaa ttagcctggg ctattatgat 660
                        caagtctccg atacaacact tgctgtttcg gaagggtggt aaatttctgc gcctactatt 720
                          agtttaagca ccaactctga gttagcgtta ggggcacagg aagggaagcg agcgaaggca 780
                            gcagggatta ttgatgccga aaaaattgag tttgtgtggg caaagacatc cgagaagaaa 840
                              atcaccttaa accacacgga taaagacgcg actatttccg cggatattgt cagtggcagc 900
```

```

gagggcctgg gctatatcaa tgcgctcaat ggcacgactt acttaaccgg tgataactct 960
gccttttagtg gtaaagtcaa aattgagcaa aatggcgctt tagggatcac ccaaaatata 1020
ggtacagcag agatcaacaa ccgcgggaaa ttacacctga aggctgacga tagcatgacc 1080
tttgccaata agatctcttg caacgggtaca ataagtatcg acagtgggac ggtggagttg 1140
5 accggcaata actatgcatt cagcggatat attgatgttg cttctggtgc tgtcgctgtt 1200
atctctgaag acaagaatat cggtcgtgca gagctggatg tcatgggcaa attgcaaatt 1260
aatgccaaca aagattgggt atttgataac gatcttgaag gtagaggcat tgttgaaata 1320
aacatgggga atcacgaatt ctcttcgat gagtttgctt atacagactg gttccagggt 1380
tactggcgt tccagaacac gacatttaat ctggaaaaga atgctgagtt tctgcagaaa 1440
10 ggcgggatca ctgcgggtca ggaagcctg gtaacagtgg gtaagggcgc tcaactccatt 1500
agcacttttg gattctccgg cggaaccgtt gattttggtg ccctgacagc aggtgcacag 1560
atgacagaag ggacgggtcaa cgttagtaaa acgctggatt tgcgcggcga ggggtgtgatt 1620
caggtttctg acagtgacgt tgtccgctca gtatctctg atattgactc tgcgttatcg 1680
ctcactgaag tcgatgatgg taacagcacc attaatgttg ttgatgcgca aggtgcggaa 1740
15 gttctggggc atgcgggcaa tctgcaattg caggataaaa atgggcaaat cctctccagc 1800
agcgcaccaac gtgatattca gcagaatggg caaaaagcgg ccgtcggcac ttacgactat 1860
cgtctgacga gtggggtaaa caatgacggg ctgtatatgt gttacggcct gaccagctt 1920
gatttacacg ctaccgacag cgatgctctg gtgctgagct ctaacggtaa aagcgagaat 1980
gccgccgatc tcagcgcaaa gattaccggc agtggtgacc tggcattcag cagccagaag 2040
20 ggtcagaccg tatcgctttc taacaaagac aacgactata ccggtgttac cgatctgcgc 2100
agtgggacgc ttttggtgaa taacgataac gtggtgggta ataccatga actgcgtctg 2160
gcggcagaga ctgaactgga catgaatggg cacagccaga ccgtgggcac gctcaatggc 2220
agcgcggatt cactgctgag cttaaattggc ggcagtctga cggttacca cgggggcaact 2280
tcaaccgggt cgtaaacggg gagcggagag ctgaatatcc agggcggcac gctggacatc 2340
25 gcgggcgata acagcaacct gacggcgaat gtgaacattg ctaattcggc taatgtcctg 2400
gtaagtcatg cgcagggtt ggtagcgca aacgttgaga acaacggta cctggcgttg 2460
aataatagcg ctgaaaaaag agcggctgcg tctgtgaatt acgccctggg cggcaatctg 2520
accaacaacg gtacgctgat gaccggaatg tcaggacagc aagctggcaa tgtgttagtg 2580
30 gtgaagggga actaccacgg taataacggg caactagtaa tgaatacggg actgaatggc 2640
gatgactcag taaccgataa attggttgtc gagggcgata ctacggcac gactgccgtt 2700
acggtgaata acgctggcgg tacaggtgcg aaaaccctta acggtatcga acttatccat 2760
gtagacggta agtctgaggg cgaatttgtt caggttgggc gtatcgttgc gggggcgat 2820
gactacactc tcgcgcgttg acaaggggca aatagtggta actggtatct gaccagcggc 2880
35 agtgattctc ctgaactgca gccggagcca gaccgatgc cgaatccaga gccaaaccgg 2940
aatccagagc cgaaccctaa ccgacacct acgcccgggc cggatctgaa tgtggataat 3000
gacctgcgac cggaggcggg tagctacatt gcgaacctg cagcagcgaa taccatgttc 3060
accacgcgctc tgcattgagc tctgggtaat acgtactata ccgacatggg gacgggtgag 3120
cagaaacaaa ccactatgtg gatgcgccat gaaggtgggc ataataaatg gcgtgatggc 3180
40 agcggccagc tgaaaaccca aagcaatcgc tatgttctgc aactgggagg cgatgtcgcg 3240
cagtggagcc aaacggcag cgaccgctgg catgttgggg tcatggcggg atatggcaac 3300
agcgacagca aaaccatttc ctgcgcaacc gggtatcgtg caaaagcgag tgtgaacgga 3360
tatagcacag gcctctatgc cacctgggtat gccgatgacg agtcgcgtaa tggcgcgtat 3420
ctcgacagtt gggcgagta cagctgggtt gataacacag tgaaagggga tgacttacia 3480
45 agtgaatcct ataaatcaaa aggatattacc gcttactagg aagctggata caaacacaaa 3540
ttagctgaat ttaatggcag ccagggaaag cgtaatgaat ggtatgttca gccgcaagca 3600
caggttacct ggatgggagt caaagccgat aagcaccgag aaagcaacgg aaccctcgtt 3660
catagcaacg gtgatggcaa tgttcaaacc cgacttggcg taaaaacctg gctgaagagc 3720
caccataaaa tggatgacgg taatccccgc gagttccagc cgtttgtaga agtgaactgg 3780
ctacataaca gtaaggattt cagcaccagt atggatggcg tgtctgtcac tcaggatgga 3840
50 gcccgaataa ttgctgagat aaaaaccggg gtggaaggac agctaaatgc caacctgaat 3900
gtctggggga atgtgggctg tcaggttgcc gataggggat ataatgacac ctctgcaatg 3960
gttggcatta agtggcaatt ctga 3984

```

<212> Type : DNA

<211> Length : 3984

55 SequenceName : SEQ ID 397
SequenceDescription :

Sequence

60 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

atgataacga tgaaaaaaag tgtattgacg gcgtttataa ctgtgggtatg tgcaacgtcc 60
agcgttatgg ctgctgatga taatgctatc acggatggct cagtaacatt taatggtaaa 120
gttattgctc cagcttgtag cctggtagct gcgacgaaag attccgtggg gactttgcca 180
65 gatgttagtg ccacgaagtt gcaaaccaat ggtcaggttt ctggcgtgca aactgatgtg 240
ccaattgaat taaaagattg tgatactacc gtaacaaaaa atgcaacgtt cacctttaat 300
ggcactgcgg atactactca gattacagcg tttgctaacc aggcctcatc tgatgctgct 360

```


acaaacgtgg ccctgcaaat gtatatgaat gatggtacaa cggccatcaa gccagacaca 420
gaaaccggga acattttgtt gcaagatgga gatcagacgt tgacttttaa agttgattat 480
atcgctacgg ggaaagcgac ttctggtaat gtgaatgcgg taacaaattt ccatattaac 540
tattattaa 549

5 <212> Type : DNA
<211> Length : 549
SequenceName : SEQ ID 398
SequenceDescription :

10 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgagtaagt ttgtaaaaac agctattgct gcaacaatgg taatgggtgc gttcgcttct 60
15 acttcaacaa tcgccgctgg caacaatggg acagcacgtt tctacggcac cattgaagat 120
tccccgtgct ctatcgttcc tgatgatcac aaactggaag ttgatatggg tgacattggg 180
tcagggatcc tgaaaaataa cgggacttct acaccgaaag ctttccagat ccatctgcaa 240
gactgtgtgt ttgacaccca gacaacgatg accactacct tcaccggtaa cgcgtcttct 300
accaacagcg gcaattacta caccatttac aataccgata ctggtgcggc atttaacaat 360
20 gtcagcctgg ccattgggtga cgctcaggga acctcttata aaagcggcgc gggatcga 420
cagaaaatcg taaacgatac ggcgaccaac aaaggcaaag cgaagcagac gctggacttt 480
aaagcctggc tgggtggcgc tgctgatgcg ccagatttag gtaattttga agccaacacg 540
accttccaga ttacttatct ctaa 564
<212> Type : DNA
25 <211> Length : 564
SequenceName : SEQ ID 399
SequenceDescription :

30 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgagggtta tctttctacg caaggagtat ttatctctac tcccgtaaat gattgcatct 60
cttttctctg ctaacggtgt cgcggcgccg attgatttat gccagggata tgatatcaaa 120
35 gcgagttgtc acgccagcag gcaaaagcctt tcaggcatta cgcaggctctg gagtattgcc 180
gatgggcaat ggctggtttt ttccgatatg accaataatg ccagcgggtgg ggccgtattt 240
ttgcaacaag gagcgggaatt tacattatca ccagaaaatg aaactggaat gactctgttt 300
gccaataaca ccgtttcagg agaataaat aacggcgggg caatatttgc taaagaaaac 360
tcaacgctga atcttacgga tgttattttt tctggtaacg tcgcaggcgg ctatgggtggc 420
40 gcaatctatt cttctggtac taacgatacc ggtgccatcg atttacgtgt cactaacgcc 480
gtgtttcgca ataacatcgc taatgacggc aaagggtggtg caatttatac catcaataat 540
gatattctatt taagtgatga tgtttttaac aataaccagg catatacatc aacaagttac 600
agtgatggcg atggcggcgc aatcgatgtc acagataata atagcgcagc caagcatcct 660
tcagggtata cgataataaa taacactgcc tttacaaata acactgccga aggttatggc 720
45 ggggcgatat ataccaatag cgcgacggct ccctatctta ttgatatttc tgttgatgac 780
agctacagcc agaacggagg cgtgttagtc gatgagaaca atagcgcagc aggttatgga 840
gatggctcct cctctgcggc ggggtggcttt atgtatctcg gcttaagtga agttaccttt 900
gatattgccg acggaaaaac gctggttatt ggcaatacag agaatgacgg agctgttgac 960
tctattgctg gtaccgggtt aatcaccaaa acagggtccg gcgatctggt acttaatgca 1020
50 gataacaatg actttactgg tgagatgcag attgaaaacg gtgaagttac cctgggcccgc 1080
agcaactccc tgatgaatgt cggcgatacg cattgccagg acgatccgca agactgctac 1140
ggctctgacga tagggagtat tgataagtac cagaatcagg cagagctaaa tgttggtccc 1200
accaacaaa cctttgcgca ctcatcgacg ggctttcaga atggcacttt aaatatcgat 1260
gctgggtggca atgttactgt taatcaaggc agttttgctg gcaccatcga aggtgctggg 1320
55 cagctcacca ttgcgcaaaa cggcagctat gtgctggcgg gggcgagtc gatggcgcta 1380
accggcgata tagtgggtgga tgctgggtgc gtgctttcgc tgggaaggcga cgcggcagat 1440
cttgccgctc tccaggacga tccgcagtcg atcgtgttaa acggcggtat gctcgatctc 1500
tctgatttct ccacctggca gagcgggtaca tcatacaaag atggccttga agtcagtggc 1560
agcagcggaa cggttatcgg cagtcaggat gtggtagatc ttgcaggcgg aaacgatatg 1620
60 catatcggcg gcgacgggaa agatggcgctc tacgtgggtga tcgatgcggg tgacgggcag 1680
gtcagcctgg caaatgacaa tcaataacctc ggcacaacgc aaatcgcttc cggtagcctg 1740
atggtgagcg acaactcgca gcttgatata acccattata accgccaggg tatctttacc 1800
gataagccac aagaaagcgt gatggagatt actgccaatg tcgatactcg ctctacaacg 1860
actgagcatg ggcgtgatat tgaaatgcgc gccgacgggtg aagtggcagt tgatgcgggg 1920
65 gtagacacgc agtggggcgc actgatggct gacagcagcg ggcagcatca ggatgagggt 1980
agcacattga ctaaaacggg ggcgggtaca ctggagctga ccgccagcgg tacaacgcag 2040
tcagcgggtga gagtagaaga gggcacgctg caagggtgatg ttgcggatat cttcccttat 2100

| | | | | | | | |
|----|---|-------------|-------------|-------------|-------------|------------|------|
| | gcttcgctgc | tatgggtcgg | tgacggggca | acgttcgtta | ctggcgcgga | tcaggatatt | 2160 |
| | cagtcaattg | atgctacttc | cagcggcact | atcgacatca | gcgatgggtac | ggttttgctc | 2220 |
| | ctgaccgggc | aggatacttc | cgctcgccctt | aatgcctcac | tgtttaactg | cgatgggacg | 2280 |
| | ctgggtgaatg | ccaccgatgg | tgtgacgttg | acaggtgagc | ttaataccaa | ccttgaaact | 2340 |
| 5 | gacagcctga | cttatctttc | caacgtgacg | gttaatggca | atctgaccaa | tacgtccggt | 2400 |
| | gcgggttagcc | tgcaaaatgg | cgctcgctggc | gatacgctga | cggtaaaccg | tgattatacc | 2460 |
| | ggcggcggtg | cgctactgct | cgatagcgaa | ttaaaccggc | atgactcggg | aagcgatcaa | 2520 |
| | ttggtgatga | acggtaatac | tgctggcaac | acaactgtgg | tggttaactc | cattacaggg | 2580 |
| | attggtgagc | cgacatcgac | aggcattaaa | gtgggttgatt | tcgcagctga | tcccacgcag | 2640 |
| 10 | tttcaaaaaca | atgcgcagtt | cagtctggca | ggcagcggct | acgtcaatat | gggagcgtat | 2700 |
| | gactacacgc | tggtggaaga | taacaacgac | tggtatctgc | gatcgcaaga | agtaacgccg | 2760 |
| | ccatcgccac | ctgatccaga | cccgactccc | gacctgatc | ccacgcagga | tcctgatcca | 2820 |
| | acaccgcacc | cggaaacctac | gcctgcttac | cagccgggtg | tgaatgcaa | agttggcggt | 2880 |
| | tatctcaata | acctgcgggc | ggcaaatacag | gcgtttatga | tgagcgacg | cgatcacgca | 2940 |
| 15 | ggtggcgatg | gtcagacgct | gaatttacgt | gttatcggcg | gagattatca | ttacacagca | 3000 |
| | gcggggcaac | tggtctcaaca | tgaagacact | tctacgggtg | aacttagcgg | cgatctgttt | 3060 |
| | agcgggcgct | ggggcacgga | tggtcgagtgg | atgcttggga | ttgttggtgg | ctacagcgat | 3120 |
| | aaccagggcg | acagccgctc | gagtatgacc | ggaactcgcg | ccgataacca | gaaccacggt | 3180 |
| | tatgcggttg | ggctgacctc | aagctgggtt | cagcacggta | agcagaagca | gggggcctgg | 3240 |
| 20 | ctggataact | ggttgcagta | cgctgggtt | agcaatgatg | tttctgaaca | tgaagatggc | 3300 |
| | gtggatcatt | accattcgct | ggggattatc | gcctcgctgg | aagcggggta | tcagtgggtt | 3360 |
| | ccggggcgctg | gtgtgggtgat | tgaaccgcag | gcgcaggtga | tttatcaggg | cgtgcagcag | 3420 |
| | gatgatttta | ccgccgctaa | ccgtgcgcgc | gtgtcacaat | cgcaggggtga | tgatattcag | 3480 |
| | acgcggctgg | gtttacacag | cgaatggcgt | accgctgttc | atgtcatacc | aacattagat | 3540 |
| 25 | ctgaattatt | atcacgatcc | ccattcgacg | gaaattgaag | aagatgccag | cactatcagt | 3600 |
| | gacgatgcgg | tgaagcaacg | gggtgaaata | aaagtgggag | tcacgggcaa | tatcagtcag | 3660 |
| | cgagtttcgc | tgctgggtag | cgtggcggtg | cagaaagggg | gtgatgattt | tgcccagacg | 3720 |
| | gcagggtttt | tgtcgatgac | ggtgaaatgg | taa | | | 3753 |
| | <212> Type : DNA | | | | | | |
| 30 | <211> Length : 3753 | | | | | | |
| | SequenceName : SEQ ID 400 | | | | | | |
| | SequenceDescription : | | | | | | |
| | Sequence | | | | | | |
| 35 | ----- | | | | | | |
| | <213> OrganismName : Escherichia coli O157:H7 | | | | | | |
| | <400> PreSequenceString : | | | | | | |
| | atgcactcct | ggaaaaagaa | acttatagta | tcacaattag | cattggcttg | cactctggct | 60 |
| | atcacctctc | aggctaattg | agcgaccaac | gatattttctg | gtcaaactta | caatactttc | 120 |
| 40 | catcactaca | acgacgccac | ctatgctgat | gacgtttact | atgatgggtta | tgtaggctgg | 180 |
| | aacaactatg | ccgctgatag | ctattacaac | ggcgatatct | acccggtcac | taataacgct | 240 |
| | accgttaacg | gcgtgatttc | tacctactat | ctggacgacg | gtatttctac | caatacgaac | 300 |
| | gccaatagtc | tgacaatcaa | aaacagcact | attcacggta | tgattacctc | tgagtgcacg | 360 |
| | actactgatt | gtgctgatga | ccgtgctact | ggttatgttt | atgatcgtct | gacactgagc | 420 |
| 45 | gttgataatt | caacgatcga | tgacaactac | gagcattata | cttacaacgg | tacctataat | 480 |
| | aatgccgctg | acactcatgt | tgtagatgtt | tacgatatgg | gtactgctat | tacactggat | 540 |
| | caggaagttg | atctgtccat | cactaataac | tctcatgtag | caggtattac | gctgactcag | 600 |
| | ggttatgagt | gggaagatat | tgacgacaac | acagtcagca | ctggcgtaaa | cagcagcgaa | 660 |
| | gtgtttaata | acactattac | tggttaaagat | tctactgtga | cctctgggtc | atggactgat | 720 |
| 50 | gaaggtacta | ctgggttggtt | tggtccatact | ggtaatgcca | gcaactatag | caacacgctg | 780 |
| | actgcagacg | atggttgcaat | tgccgcaatc | gcaaatccgt | atgctgataa | tgcatgacg | 840 |
| | actacagtaa | cttttagacaa | ctcaacactg | atgggtgatg | ttgttttctc | cagtaatttc | 900 |
| | gatgaaaact | tcttcccgcg | agggtgctaac | agctatcgcg | atgctgatgg | tgatgtagat | 960 |
| | accaacgggt | gggatggcac | agaccgtatg | gatgtgactc | tgaacaacgg | cagcaagtgg | 1020 |
| 55 | ggtggcgctg | caatgtctgt | tcatatgggt | gatgaagatg | gtgatgggtc | ttacgacgga | 1080 |
| | tatgctgttg | gtactgaagc | aactgcaact | ctgctcgata | ttgcagctaa | cagcctgtgg | 1140 |
| | ccttcatcaa | ctgtcggtgt | tgataacatc | aatactcaat | atgacgaaaa | tggtccatct | 1200 |
| | gtaggaaacg | aagtttacca | gagcgggttg | tttaattgtga | ctttgaacgg | tggttcagag | 1260 |
| | tggtgatacaa | caaaatcttc | tctgattgat | actttaagta | ttaacagcgg | ttcccaagtt | 1320 |
| 60 | aatggttcag | actctcgtct | gatctctgac | actgtctctc | tgactggcgg | ttctaacctg | 1380 |
| | aacatcggtg | aagacgggtca | tgtagcgact | aataccctga | ccatcgacaa | tagtaccggt | 1440 |
| | aaaatgtctg | atgatgtttc | tgccgggctgg | ggtttagaag | atgctgcact | gtacgcaaat | 1500 |
| | accatcaccg | taactaacga | cggctctgttg | gatattaacg | ttgatcagtt | cgatgctaac | 1560 |
| | ccgttccagg | ccgataccct | gaatctgacc | agtaccactg | ataactaacg | caacattcac | 1620 |
| 65 | gctgggtgat | tcgatatacca | tagcagtgat | tacgtaatgg | ataccgatct | ggtcaacgat | 1680 |
| | cgtaccaacg | atactaccaa | gtcaaactac | ggttatggct | taatcgcaat | gaactctgat | 1740 |
| | ggtcacctga | ctattaacgg | taacggcgat | aacgacaaca | ctgcttctat | cgaagctggg | 1800 |

cagaacgaag ttgataacaa cgggtgaccat gttgcagccg cgaccggtaa ctacaaagtt 1860
cgtatcgaca acgctactgg tgctgggtct atcgctgact acaacggcaa cgagctgac 1920
tacgtcaacg acaaaaacag caacgcgacc ttctctgctg ctaacaaagc tgacctgggt 1980
gcatacacct atcaggctga acagcgcggt aacaccgttg ttctgcaaca gatggagttg 2040
5 accgactacg ctaacatggc gctgagcatc ccatctgcga acaccaatat ctggaacctg 2100
gaacaagaca ccgttggtac tcgtttgacc aactctctgc atggcctggc tgataacggc 2160
ggcgcatggg taagctactt cgggtggtaac ttcaacggcg acaacggcac catcaactat 2220
gatcaggatg ttaacggcat catggtcggg gttgatacca aaattgacgg taacaacgct 2280
aagtggatcg tcggtgcggc tgcaggcttc gctaaagggt acatgaatga ccgttctggt 2340
10 cagggtggatc aagacagcca gactgcctac atctactctt ctgctcactt cgcgaacaac 2400
gtctttgttg atggtagctt gagctactct cacttcaaca acgacctgtc tgcaacctatg 2460
agcaacggta cttacgttga cggtagcacc aactccgacg cttgggggctt cggtttgaaa 2520
gccggttacg acttcaaact ggggtgatgt gggtacgtga ctcttaacgg cagcatttct 2580
ggctctgttc agtctggtga tgactaccag ctgagcaacg acatgaaagt tgacggtcag 2640
15 tcttacgaca gcatgcgtta tgaactgggt gtagatgcag gttatacctt cacctacagc 2700
gaagatcagg ctctgactcc gtacttcaaa ctgggttacg tctacgacga ctctaacaac 2760
gataacgatg tgaacgggtga ttccatcgat aacgggtactg aagggtctgc ggtacgtgtt 2820
ggctctgggtg ctacgttcag cttcaccaag aacttcagcg cctataccga tgctaactac 2880
ctcgggtggtg gtgacgtaga tcaagactgg tccgcgaacg tgggtgttaa atatacctgg 2940
20 taa 2943
<212> Type : DNA
<211> Length : 2943
SequenceName : SEQ ID 401
SequenceDescription :
25 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
30 atgaaactca aacatggttg tatgattgtc gtttctgtgt tggcgatgtc gtctgctgctg 60
gtaagcgcag ccgaggggtga tgaatcagta acgaccactg ttaatggcgg tggtattcat 120
tttaaagggtg aagtggtaaa tgccgcttgt gcgattgatt ccgaatcaat gaaccaaacg 180
gttgagctgg gtcagggttcg ttcttctcgc ctggctaaag cgggtgacct cagctccgcc 240
gttggtctca atatcaagct gaatgattgt gataccaatg ttccagtaa tgcagctgtt 300
35 gcattcctgg gtactactgt caccagtaat gacgatacgt tagcgtgca gagttcagcg 360
gcaggctctg ccaaaaatgt cggatttcaa attttggacc gtacgggtga ggtattaata 420
cttgatgggg ccacttttag tgctaaaacc gacttgattg atggcacgaa tatactacca 480
ttccaggctc gttatattgc tctcgggcag tccgtagctg gtactgcaaa cgcagatgctg 540
accttcaaag ttcaatatct ataa 564
40 <212> Type : DNA
<211> Length : 564
SequenceName : SEQ ID 402
SequenceDescription :
45 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 atgaaacttt taaaagtagc agcaattgca gcaatcgtat tctccggtag cgctctggca 60
ggtgttggtc ctacgtacgg cggcggtggc ggtaaccacg gtggtggcgg taataacagc 120
ggcccgaatt cagagctgaa tatttatcag tacgggtggg gtaactctgc acttgctctg 180
caagctgatg ctcgtaactc tgatcttact attaccagc atgggtggtg taacggtgca 240
gatgttggtc agggctcaga tgacagctca atcgatctga cccaacgtgg ctttggtaac 300
agcgccactc ttgatcagtg gaacggtaaa gactctcata tgacagttaa acaattcggt 360
55 ggcggcaacg gtgcagcggg tgaccagact gcattcaatt ccaccgtcaa cgtaactcag 420
gttggttttg gtaacaacgc gaccgctcat cagtactaa 459
<212> Type : DNA
<211> Length : 459
SequenceName : SEQ ID 403
SequenceDescription :
60 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
65 atgcctattg gtaatcttgg tcataatccc aatgtgaata attcaattcc tcttgcacct 60
ccattacctt cacaaccga cgggtgcagg gggcgtgggt agctcattaa ctctacgggg 120

ccgttgggat ctcgtgcgct atttacgcct gtaaggaatt ctatggctga ttctggcgac 180
aatcgtgcc a gtgatgttcc tggacttcct gtaaataccga tgcgcctggc ggcgtctgag 240
ataacactga atgatggatt tgaagttctt catgatcatg gtccgctcga tactcttaac 300
aggcagattg gctcttcggt atttcgagtt gaaactcagg aagatggtaa acatattgct 360
5 gtcggtcaga ggaatgggtg tgagacctct gttgttttaa gtgatcaaga gtacgctcgc 420
ttgcagtgca ttgatcctga aggtaaagac aaatttgtat ttactggagg ccgtgggtgg 480
gctgggcatg ctatggtcac cgttgcttca gatatacagg aagcccgcca aaggatactg 540
gagctgttag agcccaaagg gaccggggag tccaaagggt ctggggagtc aaaaggcgtt 600
ggggagttga gggagtcaaa tagcgggtgcg gaaaacacca cagaaactca gacctcaacc 660
10 tcaacttcca gccttcgttc agatcctaaa ctttgggttg cgttggggac tgttgctaca 720
ggctctgatag ggttggcggc gacgggtatt gtacaggcgc ttgcattgac gccggagccg 780
gatagcccaa ccacgaccga cctgatgca gctgcaagtg caactgaaac tgcgacaaga 840
gatcagttaa cgaaagaagc gttccagaac ccagataatc aaaaagttaa tatcgatgag 900
ctcggaaatg cgattccgtc aggggtattg aaagatgatg ttgttgcgaa tatagaagag 960
15 caggctaaag cagcaggcga agaggccaaa cagcaagcca ttgaaaataa tgctcaggcg 1020
caaaaaaat atgatgaaca acaagctaaa cgccaggagg agctgaaagt ttcacggggg 1080
gctggctacg gtcttagtgg cgcattgatt cttgggtggg gaattgggtg tgccgtcacc 1140
gctgcgcttc atcgaaaaaa tcagccggta gaacaaacaa caacaacaac tactacaact 1200
acaactacaa gcgcacgtac ggtagagaat aagcctgcaa ataatacacc tgcacagggc 1260
20 aatgtagata ccctgggtc agaagatacc atggagagca gacgtagctc gatggctagc 1320
acctcgtcga ctttctttga cacttccagc atagggaccg tgcagaatcc gtatgctgat 1380
gttaaaacat cgctgcatga ttgcgaggtg ccgacttcta attctaatac gtctgttcag 1440
aatatgggga atacagattc tgttgtatat agcaccattc aacatcctcc ccgggatact 1500
actgataacg gcgcacggtt attaggaaat ccaagtgcgg ggattcaaag cacttatgcg 1560
25 cgtctggcgc taagtgggtg attacgccat gacatgggag gattaacggg ggggagtaat 1620
agcgtgtga atacttcgaa taaccaccca gcgcgggat cccatcgttt cgtctaa 1677

<212> Type : DNA

<211> Length : 1677

30 SequenceName : SEQ ID 404

SequenceDescription :

Sequence

35 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

atgttttcta ctttcaaaaa agcagctctg ctggcagcta ttgcattacc tttttcaact 60
atggctgctc ctacagtcac ttttcagggt gaagtaaccg atcagacctg ttccgtaaact 120
atcaacggtc aaaccaattc agtagtattg atgccgaccg tagccatggc tgacttcggt 180
40 gcaacttttag ctgatggta gagcgcaggc cagacgcctg ttacggtttc tgtgtctaac 240
tgccaggctc caactggtgc agatcaggca atcaacacca ctttcctggg ctacgacgtt 300
gacgctagca cgggtgttat gggaaaccgt gataccagca gcgatgcggc gaaaggcttt 360
ggcattcagt taatggattc cagcacttct ggtaaccag taactctggc tggcgcgact 420
aacgtaccgg gtctgacct gaaagttggc gataccgaag ccagctacga cttcgggtgcg 480
45 cgttacttcg ttatcgatag cgctgctgcc actgccggta aaattaccgc tgtcgcagaa 540
tacaccctga gctacctcta a 561

<212> Type : DNA

<211> Length : 561

50 SequenceName : SEQ ID 405

SequenceDescription :

Sequence

55 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

atgaacagtg aaggaggaaa accggggaat gtactgaccg ttaacggcaa ctataccgga 60
aacaatggcc tgatgacgtt caacgcgacg ctgggcggcg ataattcacc caccgataag 120
atgaacgtga aaggcgatac ccaagggaac actcgcgttc gggttgataa cattggcggc 180
gtcgggtgcg aaacgggtcaa cggattgaa ctcatgagg ttggcggtaa ttctgcagg 240
60 aatttcgcgc tgaccaccgg aactgtcgaa gctggggctt acgtctacac gctggctaaa 300
gggaagggga atgacgagaa aaactgggtat ctgaccagta aatgggacgg cgtaacgcca 360
gcgatacac ccgatcccat caataatccc cctgttgttg atccggaagg cccatcagtt 420
tatcgcccgg aggcgggaag ctatatcagc aacattgccg cagccaactc gctgtttagc 480
catcgcttac acgaccgtct gggtagacgg caatatacag attcactgca ttctcaggat 540
65 tcagcaagca gtatgtggat gcgtcatgtc ggggggcacg aacgttccag tgccggagac 600
ggccagctaa atactcaggc taaccgctat gtattgcagc taggcggcga tttggcgacg 660
tgagtagca acgcgcagga tcgctggcat cttggcgtga tggcaggcta cgccaatcag 720


```

5  cacagtaata ctcagagtaa tcgtgtgggt tataaatcgg atgggcgcat cagcggttac 780
   agcgctgggc tgtacgcgac ctggtatcag aacgatgcga ataagaccgg cgcttatgtt 840
   gacagctggg cgctgtataa ctggtttgat aacagcgta gttccgataa ccgttctgct 900
   gacgactatg attctcgcgg tgtgacggcc tctgttgagg gtgggtatac ctttgaagcg 960
10 ggaacatgta gcggcagcga agggacgctg aatacctggt acgtccagcc acaggcgcaa 1020
   atcacctgga tgggtgtgaa agattctgac catgcccgga aagacggaac gcgcattgaa 1080
   acggaaggcg acggaacgt gcaaacgcga cttggggtga aaacctacct gaatagccat 1140
   caccagcgtg acgatggtaa acagcgtgag ttccagcctt acattgaagc gaactggatc 1200
   aacaatagca aagtctacgc cgtgaagatg aatgggtcaa ccgtaagccg tgatgggtgcg 1260
15 cgaaatctcg gtgaagtacg taccgggggt gaggcgaaag taaataacaa ccttagcctg 1320
   tgggggaatg tcggtgtgca actaggtgat aaaggctata gcgatactca gggcatgctg 1380
   ggagtgaat atagctggta a
   <212> Type : DNA
   <211> Length : 1401
15   SequenceName : SEQ ID 406
   SequenceDescription :

Sequence
-----
20 <213> OrganismName : Escherichia coli O157:H7
   <400> PreSequenceString :
   atgtcatatc tgaatttaag actttaccag cgaaacacac aatgcttgca tattcgtaag 60
   catcgtttgg ctgggttttt tgtccggctc tttgtgcct gtgcttttgc cgtacaggca 120
   cctttgtcat ctgccgaact ctattttaat ccgcgctttt tagcggatga tccccaggct 180
25 gtggccgatt tatcgcgttt tgaaaatggg caagaattac cgccagggac gtatcgcgtc 240
   gatattctatt tgaataatgg ttatatggca acgcgtgatg tcacatttaa tacgggcgac 300
   agtgaacaag ggattgttcc ctgcctgaca cgcgcgcaac tcgccagtat ggggctgaat 360
   acggcttctg tcgccggtat gaatctgctg gcggatgatg cctgtgtgcc attaaccaca 420
   atggtccagg acgctactgc gcatttagat gttggtcagc agcgactgaa cctgacgac 480
30 cctcaggcat ttatgagtaa tcgcgcgcgt gggttatatc ctctgagtt atgggatccc 540
   ggtattaatg ccggattgct caattataat ttcagcggaa atagtgtaca gaatcggatt 600
   gggggtaaca gccattatgc atatttaaac ctacagagtg gggttaaata tgggtgcgtgg 660
   cgtttacgcg acaataccac ctggagttat aacagtagcg acagatcatc aggtagcaaa 720
   aataaatggc agcatatcaa tacctggcct gagcgagaca taataccgtt acgttcccgg 780
35 ctgacgctgg gtgatggtta tactcagggt gatattttcg atgggtattaa ctttcgcggc 840
   gcacaattgg cctcagatga caatatgtta cccgatagcc aaagaggatt tgccccgggtg 900
   atccacggta ttgctcgtgg tactgcacag gtcactatta acaaaaatgg gtatgacatt 960
   tataatagta cggtgccgcc ggggcctttt accatcaacg atatctatgc cgcaggtaat 1020
   agtgggtgact tgcaggtaac gattaaagag gctgacggca gcacgcagat ttttaccgta 1080
40 ccctattcgt cagtcccgtt tttgcaacgt gaagggcata ctcgttatc cattacggca 1140
   ggagaatacc gtagtggaaa tgcgcaacag gaaaaacccc gctttttcca aagtacatta 1200
   ctccacggcc ttccagctgg ctggacaata tatggtggaa cgcaactggc agatcgttat 1260
   cgtgctttta attttgggtat cgggaaaaat atgggggcac tgggcgctct gtctgtggat 1320
   atgactcagg ctaattccac acttcccgat gacagtcagc atgacggaca atcgggtcgt 1380
45 tttctctata acaaatcgct caatgagtca ggcacgaata ttcagttagt gggttaccgt 1440
   tattcgacca gcggatatat taatttcgct gatacaacat acagtcgaat gaatggctac 1500
   aacatcgaaa cacaggacgg agttattcag gttaaagcca aattcaccga ctattacaac 1560
   ctcgcttata acaaacgcgg gaaattacaa ctcaccgtta ctcagcaact cgggcgctca 1620
   tcaacactgt atttgagtgg tagccatcaa acttattggg gaacgagtaa tgcgatgag 1680
50 caattccagg ctggattaaa tactgcgttc gaagatatca actggacgct cagctatagc 1740
   ctgacgaaaa acgcctggca aaaaggacgt gatcagatgt tagcgcgtaa cgtcaatatt 1800
   cctttcagcc actggctgcg ttctgacagt aaatctcagt ggcgacatgc cagtgccagc 1860
   tacagcatgt cacacgatct caacggctcg atgaccaatc tggctggtgt atacggtacg 1920
   ttgctggaag acaacaacct cagctatagc gtgcaaaccg gctatgccgg gggaggcgat 1980
55 ggtaatagcg gaagcacagg ctacgccacg ctgaattatc gcggtggtta cggcaatgcc 2040
   aatatcggtt acagccatag cgatgatatt aagcagctct attacggagt cagcgggtggg 2100
   gtactggctc atgccaatgg cgtaacgctg gggcagccgt taaacgatac ggtggtgctt 2160
   gttaaagcgc ctggcgcaaa agatgcaaaa gtcgaaaacc agacgggggt gcgtaccgac 2220
   tggcgcggtt atgccgtgct gccttatgcc actgaatatc gggaaaatag agtggcgctg 2280
60 gataccaata ccctggctga taacgtcgat ttagataacg cggtcgctaa cgttgttccc 2340
   actcgtgggg cgatcgtgcg agcagagttt aaagcgcgcg ttgggataaa actgctcatg 2400
   acgctaacc acaataataa gccgctgccg tttggggcga tggtgacatc agagagtagc 2460
   cagagtagcg gcattgttgc ggataatggc caggtttacc tcagcggaat gcctctagcg 2520
   ggaaaagttc aggtgaaatg gggagaagag gaaaatgctc attgtgtcgc caattatcaa 2580
65 ctgccaccag agagtcagca gcagttatta acccagctat cagctgaatg tcgttaa 2637

<212> Type : DNA
```

<211> Length : 2637
SequenceName : SEQ ID 407
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

| | | |
|----|--|-----|
| 10 | atgcagataa tctttggaga aaaatgcgtg tcattactac gactatTTTT tgcgcgcgtc | 60 |
| | ttaatgctat ggtgcgctca aaccgctgct tatagcgggc agtgtcatac cactcagggg | 120 |
| | aatccgtata ttggcgtcaa ttttggcgtt aaaaccctgg aggaagaaga aaatacgact | 180 |
| | ggggtagtaa aagacaaatt ttatcagtgg aacgaatcga atgattatta tgtttcctgt | 240 |
| | gattgcgata aagacaatgt cagaagtggc cgatgggcat tcgccgcgga ttcaccgtta | 300 |
| | gtctatttag gcgacaactg gtacaaaatt aatgactatc ttgccgcaa agttttattg | 360 |
| 15 | cagggttaaag gcagttctcc tacagcgggt cctttcgaaa acgtggggac tggggcagat | 420 |
| | acccggtggc atatttgtga ccccgcggt caacgtttag gcggccaggg agctagcggg | 480 |
| | aatagcggta gcttttccct gaaaatattg cagccgttcg ttggttcggt cgtcattcct | 540 |
| | cctatggcgc tggcgcgatt atttgaatgc tacaacatac ccgcaggtga ttcctgcacg | 600 |
| | actacaggca caccggtttt agtgtattac ctgtctggta ctatcaattc acttggctca | 660 |
| 20 | tggtccgtca atgccggaga aacaatcgag gtcgatctgg gcgacgtatt tgcggctaac | 720 |
| | tttcgtgttg tagggcataa gcctcttggg gccagaacgg cagaacttgc aattccagtc | 780 |
| | aggtgtaaca cgggaaacgc ggggttagtt aacgtcaacc tgagtctgac ggcaaccaca | 840 |
| | gaccccgct atccccaggg gattaagacg tcacgtcctg gcgtgggcgt ggtggtgacc | 900 |
| | gatagccaga acaacattat ttcccctgct ggtggaacat taccgtcttc tattcctgat | 960 |
| 25 | gatgcagaca gtatcgcgtg a | 981 |

<212> Type : DNA

<211> Length : 981

SequenceName : SEQ ID 408

SequenceDescription :

30

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

| | | |
|----|---|-----|
| 35 | atgaaaatta aaactctggc aatcgttgtt ctgtcggctc tgtccctcag ttctacagcg | 60 |
| | gctctggccg ctgccacgac ggttaatggt gggaccgttc actttaaaagg ggaagttytt | 120 |
| | aacgccgctt gcgcagttga tgcaggctct gttgatcaaa ccgttcagtt aggacaggtt | 180 |
| | cgtaccgcat cgctggcaca ggacggagca accagttctg ctgtcggttt taacattcag | 240 |
| | ctgaatgatt gcgataccaa tgttgcatct aaagccgctg ttgccttttt aggtacggtg | 300 |
| 40 | attgatgcgg gtcataccaa cgttctggct ctgcagagtt cagctgcggg tagcgcaaca | 360 |
| | aacgttgggtg tgcagatcct ggacagaacg ggtgctgcgc tgacgctgga tgggtgcgaca | 420 |
| | ttcagtgagc aaacaaccct gaataacggt actaacacca ttccgttcca ggccggttat | 480 |
| | tatgcaatcg gcgaggcaac cccgggtgct gctaatacggt atgcgacctt caagggttcag | 540 |
| | tatcaataa | 549 |

45 <212> Type : DNA

<211> Length : 549

SequenceName : SEQ ID 409

SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

| | | |
|----|--|-----|
| 55 | atgaaattaa aagtcacgc tacactgatt gctactgttg ccgtgggtgt aagctttaac | 60 |
| | agcaattttg cttctgcgag tacaacgtcc gcttctttaa ccgtaaacag taacctgact | 120 |
| | atgggtacct gcagtgtca gataatggat aatagtaata aagtgatcaa tgaagtggtc | 180 |
| | tttggaatg tttatatttc tgaactcggg gcaaaaagca aagtgaaca gtttaaaatt | 240 |
| | cgcttttagca attgctctgg ccttccccaa aacagcgccc aaatagtgtt ggcacctaat | 300 |
| | ggtatatcct gtgctggttc tcaatcgtca tcggcgggtt tttctaaca gtttactgac | 360 |
| 60 | gctagcgcag caaccagaac ggctgtggaa gtatggacta cagatacacc ggaaagcaat | 420 |
| | ggcagtagcg aattccattg tgcataaaag ataccagtgc ctgtgacgct tcccgcgcac | 480 |
| | accacaactc agccttacga ttaccggtta agtgcacgga tgaccgttgc ggaaggtaga | 540 |
| | ttggtaaccg atgtaagacc gggtaatttc cgctctccca cgactttcac gatcacttat | 600 |
| | cagtaa | 606 |

65 <212> Type : DNA

<211> Length : 606

SequenceName : SEQ ID 410

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
 ttggcatcaa cagttgagta tggtagagaca gttgatgggtg ttgtcctgga aaaagatatc 60
 cagctgggttt atgggaccgc caataatacgc aaaatcaatc ctggcgggaga acagcatatt 120
 aaagaatttg gtataagtag taatactgaa attaacggcg ggtatcagta cattgaaatg 180
 10 aatggcaccg cagaatactc agtattaaat gatgggtatc aaattgttca aatgggtggc 240
 gcggaacacc agactacgct caataatggt gtgctacagg tttatggcgc agcgaatgat 300
 cccacgatta aaggcgggag cttaatcggt gaaaaagatg ggattaccgt ccttgccgct 360
 atcgaaaagg gaggattact ggaggttaaa gaggggggat tagcgattgc ggtagatcag 420
 aaagcaggcg gtgctattaa agcaagcagc cgggtcatgg aggtattcgg aacaaaccgt 480
 15 ctcggtcagt tcgaaatcaa gaatggattt gctaacaata tgctgttggg aaacggcgga 540
 agttttgcgag ttgaagaaaa tgacttcgct tataatacta ctgtagatag tggcggctta 600
 ctggagggtta tggatggcgg gactgcaact ggctgtgata aaaaagcagg cggaaaatta 660
 attgtctcaa cgaatgcgct ggaagtgaat ggtacaaaca gtaaggcca atttagtata 720
 aaagatgggt tgtcaaaaaa ttatgaactg gatgatgggt ccgggcttat tgttatggag 780
 20 gacacgcagg ccattgacac tatcctcgat gagcatgcca ctatgcaatc gctgggaaag 840
 gatactggta cgagagtgcg ggcaaatgcg gtatatgatc tcggtcgatc agatcagaat 900
 ggaagtataa cgtattcctc taaagccatc tctgaaaata tggttatcaa caatggccgc 960
 gctaactgct gggctggcac aatgggttaac gtgtcagtcg gaggaatga tggcattctt 1020
 gaggttatga agccgcaaat aaattatgca ccgcgaatgt tgggtgggtaa ggtagtgggt 1080
 25 tctgagggcg cttctttaag aacgcattgt gccgtggata ccagcaaagc ggatgtttcg 1140
 ctgaaaata gcgcattggc catcattgct gatatacta cgacgaacca aaacaccgc 1200
 cttaacttag ccaaccttgc gatgtctggc gcaaatgtga ttatgatgga tgagtcagt 1260
 actcgttcat ctgtgacggc aagtgcggaa aatttacta cgttgaccac caataccctg 1320
 tcgggaaacg gcaattttta tatgcgtacc gatatggcga atcatcagag cgatcagctc 1380
 30 aacgtcaccg gtcaggcaac aggtgatttc aaaatattcg tgacggacac cgggtgccagc 1440
 ccggcagcag gagatagcct tacactggta acaacggcg gcgggtgatgc tgcatttacg 1500
 ttgggcaatg ccggaggcgt tgttgatata ggtacgtatg aatatacctt gctggataat 1560
 ggtaaccata gctggagtct ggcagagaat cgcgcgcaaa ttacccttc aaccactgat 1620
 gtgctgaata tggcggccgc acaaccgctg gtatttgatg cagaactgga caccgtgcgt 1680
 35 gagcgtcttg gtagcgtaaa aggcgttagt tacgatacgg cgatgtggag ttcggcaatt 1740
 aacaccgca acaacgtgac cactgatgcg ggagctgggt ttgagcaaac attgacgggc 1800
 ctgacgctcg gtatcgatag ccgtttctcc cgtgaagaaa gcagcacaat tcgcggcttg 1860
 ttctttgggt actctcattc tgatattgggt tttgatcgcg gcggcaaagg caatgtcgat 1920
 agctataccc tgggggctta tgccgggttg gagcatcaga acggtgccta tgttgatgga 1980
 40 gtggtgaaag ttgaccgttt tgccaacacc atccatggca agatgagtaa tggggcaaca 2040
 gcgtttggcg attacaatag taacggcgcg ggtgctcatg tcgagagcgg gttccgttgg 2100
 gttgacggat tgtggagtgt tagaccctat ctggccttta ccggctttac cacagatgg 2160
 caggactaca cgttatcaaa cggcatgcgc gcggatgtgg gaaatacccg gatattacgc 2220
 gctgaagcgg gaacggcggg aagctatcac atggacctgc aaaacgggtac gacgctggaa 2280
 45 ccctggctga aagccgcggt gcgtcaggaa tacgccgatt ctaaccagggt gaaagttaat 2340
 gacgatggca aatttaataa tgatgtggct ggaaccgctg gcgtttatca ggctgggata 2400
 aggtcatcgt ttaccctgac gttaagcgggt catttgtcag tcagctatgg caatggcgca 2460
 ggggtagaat cgccgtggaa taccaggcg ggtgtggtct ggacgttctg a 2511

50 <212> Type : DNA
 <211> Length : 2511
 SequenceName : SEQ ID 411
 SequenceDescription :

55 Sequence

<213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
 atgcaaagga aaggcaataa actgttgatt cagttatgca gtgtgatact gctatttttt 60
 60 accacatcct ggtatgcatt ggcgaatgaa tgttatatag agagaaatgc tgaaggggat 120
 tatcacatga agataagctc tactcagctt agtctggcgt cacaataggc cgagggtccg 180
 acagaaatag ccgaagctac atgggatgta aatattcaac taagaggcga tgccataggg 240
 tgtaaatctc ttggggatag taaggcagtt cactttctta atacagctga cccaagttta 300
 atatccacgt acaccacaac gaatggcgca gcgttattaa aaacaactgt tccaggcatt 360
 65 gtgtattctg tcgagttatt atgccttagt tgtggtgccg cagatgaact tgatttatgg 420
 ctacctgcac aaagtggcgc agataacttc ataccaagca ccagacgaa atgggcctat 480
 gagtacagt atcaaagttg gtatttacgt tttcgcttat tcataactcc tgaatttaaa 540

cccaagaatg gtgtttccag cggaacaacg atagcaggaa agattgcgtc atggtatata 600
ggtaccaatg accagccgtg gatcaacttt tacattgaca atgactcttt aaagtttttc 660
gtcgatgaac cgacctgtgc aacagttgcc ctggcacaag atcagggcaa cgtcagtggc 720
aatcaggtaa cgcttgggaa cagctatgtt tcggaagtga aaaatgggct tacgcgggaa 780
5 atcccttttt ctatccgtgc tgaatactgt tatgccagta aaattacggg taagttgaaa 840
gcggaacaata aaccagcga tgccacactg gtgggtaaaa cgaactggctc ggcttcaggc 900
gtggctgtaa aagtaaattc aacttatgac aatagcaaag tattgttaaa agcagatggg 960
agcaacacgg ttgactacaa cttcgccgcc tggcacaaca acctgctgtt ttacctttt 1020
acggcgacgc tggtagcgga tggtagcggt aatgctgtcg gtgttggaac attttcaggg 1080
10 aacgcgacct tctcctttac ctacgaataa 1110
<212> Type : DNA
<211> Length : 1110
SequenceName : SEQ ID 412
SequenceDescription :
15 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
20 ttgtaccagt ttactcatca aaaaagccgt atcccgaata aaacgctact tgcggcctgt 60
tgtgccctgt tttatagcag caacgggtgt ggcggcgaca ccgtggaata tgacagttcc 120
tttttaaatg gaactggcgc atcaacgatt gatgttaaac gttatgtcga aggcaaccgc 180
acaccgcccgt gtctctataa tgtccgcgta tttgtaaacg gtcaggcgac ttccagctta 240
gaaattccgt ttgtggatat tggcgaatac agtgccggcg cctgtcttac ccataaaaac 300
25 ctggcgcaac ttcacattaa gcaacctgaa cagcctgtca ctttactcgc cagagaaggt 360
gaagaagagg attgtctgga tctggcaaag tcatacgaag aggcggatgt gtgctttgac 420
ggtagtgacc agtttctcga tctgacgatc cctcaggcct atgttctgaa aagctatggc 480
ggctacgttg acccttcttt atgggaatcg ggaattaacg ctgccacact ggcataatcc 540
ctgaacgcgt atcacacaag ttcagataac gacaatagtg acagcgtcta tggcgcgctc 600
30 aactcaggta tcaatttagg agcctggcac tttcgtgcgc gcggtaacta taactggaca 660
acagataacg gcagcgattt cgatttccag gatcggtact tacagcgtga cattccggca 720
atccgttccc agataattat gggtagatgc tataccaccg gtgaaacgtt tgactctgtc 780
aacgtccgtg gtgttcgcct gtacagcgac agccgtatgc tgccttcggc gctggccagt 840
tacgctccga ccatccgcgg tgtagcaaac tccaacgcca aagtcaccgt gacgcaaagc 900
35 ggatataaaa tttatgaaac caccgttccg cccggtgaat ttgttataga cgacattagc 960
ccttccggct ttggtagcga actgggtcgt accattgaag aagcggatgg ttccaaacgc 1020
acctttacgc aacccttctc gtcgggtgtg caaatgcaac gtcctgggtg gggccgttgg 1080
gatttcagcg cgggtaaagt cattgatgac agtctgcgat ccgaacccaa tatggggcaa 1140
gcctcttatt actatgggtc gaataacctc ttcacgggtt ataccggcat tcagttcacc 1200
40 gataataact atcttgccgg gctgttaggt gtgggtatca acaccagcat cggcgccctt 1260
gcggtagacg ttaccatttc ccgtgctgaa attccggatg ataaaacctc ccaggggcaa 1320
agttatcgcg tgacctgga caaacttttc caggataacc ggacatcatt taacctcgcg 1380
gcgtacgcgt attccaccca ggattacctg ggcctgcagt atgcgttagt cctcattgac 1440
gacgccaagc atttgtctgc cgatgaagac aaaaacacca tgcagacgta ctcacgtatg 1500
45 aaaaaccagt ttaccgtcag cattaaccag ccattgaata tcgcctatga agattacggg 1560
tcgctgttta tttccggtag ctggacgtat tactggggcg cgaacaatag ccgactgaa 1620
tataatgttg gttacagtaa aagcgtttcg tggggcagtt tcagcgtcaa cctacaacgt 1680
agctggaatg aagacggcga gaaagatgac gcgatgtacg tcagcgttag cgtacctatt 1740
gagaatatat taggtggcaa acgtaagtct tctgggttcc gcaattttaa tactcagctc 1800
50 aataccgatt tcgatgggtc acatcagttg aatgttaaca gttccggtaa cactgaaaac 1860
aatctgggtg actacagtgt caacgcaggt tatagcctcg ataaaaacgc cggcgattta 1920
gcctctgttg gtggttatct caactatgaa tctgggttag gcggtatttc cgcttcggcc 1980
tcggccactt ctgataacag ccaacagtac tccatctcaa ccgatggcgg ctttgtatta 2040
cacagtgggt gtttaacgtt cactaacaac agtttcagca gtaacgacac gctgggtgta 2100
55 atcaacgccc taggtgctaa aggcgcacga atcaataaca gtaataacga aatcgatcgc 2160
tggggatatt ccgtgacgtc ctctgtcagc ccatatcgtg aaaaccgggt aggtctgaac 2220
attgaaacac tggaaaacga tgttgaaactg aaaagtacca gcgccaccac cgtaccacgt 2280
agcggctccg ttgttttgac ccgtttcgaa actgacgagg ggcgttctgc cgtgctgaat 2340
attactgccg ccaatggcaa atccattccg tttgctgcgg aggtttacca gggtaggggtg 2400
60 atgatcgcca gcatgggcca ggggtggtcag gcatttgtac gcggtattaa cgacagcggg 2460
gaattaatcg tgcgtggta tgaatacaac caaaccattg actgtaagtt gcactaccag 2520
ttcccgccgc agccacaaac gcagggaagc accaacacct tattacttaa caatcttacc 2580
tgtcaggtag caaatcacta a 2601
<212> Type : DNA
65 <211> Length : 2601
SequenceName : SEQ ID 413
SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaagttca aacgattgct gcatagcggc atcgccagtt tgagtctggt tgccctgcggg 60
gtgaatgcgg cgacggatct tggcccggca ggggatattc atttctccat cactatcacc 120
actaaagctt gcgagatgga aaaaagcgat ctogaagtcg atatgggaac aatgacgctg 180
caaaaacctg cggcagtcgg tacggtgttg agcaagaaag atttcacat tgaactcaaa 240
10 gagtgcgatg ggatatccaa agcgaccgtt gagatggaca gtcagtcgga cagcgatgat 300
gattccatgt ttgcccttga ggctgggtggc gcaacgggtg ttgcgttgaa gatagaggac 360
gataaaggaa cgcagcaagt tcccaaaggc tccagcggaa cgccgattga atgggacgatt 420
gatggcgaaa ccacgtcgct tcaactaccag gcgagttatg tggtcgtcaa cactcaggcc 480
actggtggca cagcgaatgc ccttgtaaatt ttttccatca cctatgagta a 531

<212> Type : DNA

<211> Length : 531

SequenceName : SEQ ID 414

SequenceDescription :

Sequence

25 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaataca ataacattat tttcctcggt ttatgtctgg gggttaaccac ctattctgct 60
ttatccgcag atagcggtat taaaattagc gggcgcgctc tcgattatgg ctgcacagtc 120
tcatcggtat cgcttaattt taccgtagat ctccaaaaaa acagtgccag acaatttcca 180
acgaccggta gcacaagtcc agccgtccct tttcagatta cggttaagtga atgcagcaaa 240
gggacaacgg ggggttcgggt tgcatttaac ggtattgagg acgcagaaaa taataactctg 300
30 ttgaaactgg atgaggggaag caatacggcc tccgggtttag gtatagaaat actggacgga 360
aatatgcgtc cgggtgaaact gaatgacctt catgccggga tgcagtggtat cccactggta 420
ccagaacaga acaatatattt gccttactcc gctcgtctga agtcaactca gaagtccgtc 480
aatccgggac tgggtgagggc ttcggcaacc tttacccttg aatttcaata a 531

<212> Type : DNA

<211> Length : 531

SequenceName : SEQ ID 415

SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaatggc gcaaacgtgg gtattttattg gcggaatat tggcgctcgc aagtgcgacg 60
45 atacaggcag ccgatgtcac catcacggtg aacggtaagg tcgtcgccaa accgtgcaca 120
gtttccacca ccaatgccac ggttgatctc ggcgatcttt attctttcag tctgatgtct 180
gccggggcgg catcggcctg gcatgatgtt gcgcttgagt tgactaattg tccgggtggga 240
acgtcaaggg tcaactgccag cttcagcggg gcagccgaca gtaccggata ttataaaaac 300
caggggaccg cgcaaaacat ccagtttagag ctacaggatg acagtggcaa cacattgaat 360
50 actggcgcaa ccaaaacagt tcaggtggat gattcctcac aatcagcgca cttcccgtta 420
caggtcagag cattgacggg aaatggcgga gccactcagg gaaccattca ggcagtgatt 480
agcatcacct atacctacag ctga 504

<212> Type : DNA

<211> Length : 504

SequenceName : SEQ ID 416

SequenceDescription :

Sequence

60 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaagag cgcctcttat aacaggactt ttgttgatat ccacatcctg cgcttatgcc 60
tcctcagaag ggtgtggagc tgacagcact agcgggtgca caaattacag cagtgtgggt 120
gatgatgtta cgggtgaacca gacagataac gtgacaggac gggagtttac ctctgcaacg 180
65 ctaagtagca ctaactggca atacgcctgt tcctgctctg cgggtaaggc agttaaactt 240
gtctatatgg tcagccccgt acttaccacc actggacatc agacaggata ttacaaactc 300
aatgacagcc tggatattaa aaccatgaac cgccccggaa atcctggaga ctaa 354

<212> Type : DNA
<211> Length : 354
SequenceName : SEQ ID 417
5 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
10 <400> PreSequenceString :
atgaaaaaag cacttctcgc agccgctctg gttatggctt ctggttccgc cctggctgta 60
gatgggtggc atatcgactt taacgggatg gtacagtccg gtacctgtaa agtgggtgtg 120
gtagatactg gtatgcatag cgttaccact gatggcggtg ttaccctgga tactgcgaat 180
gttactgata cttttgctga agtttagcga actgctgtcg gtttactgcc gaaagagtgc 240
15 atgatttctg ttgagtgtga tccagggtgct ccgaagaatg ctgagttaac tatgggttct 300
gcaagttacg cgaacaccag cgggtaccctg aataacaata tgaacatcac tgttaacggg 360
attgcaccgg ctcagaacgt aaacattgca gttcataaca tgaaaaacaa agctggcgct 420
gctgaaatta agcaggtcca tatgaacaac tcttctgaag ttcaggaact gacattagac 480
gcagaaggta aaggccagta cgtatttaac gcatcttacg ttaaagcacc gaacagcccg 540
20 gctgtaactg ctggtcatgt aaccactaac gcgctgtaca ccgttgctta taagtaa 597

<212> Type : DNA
<211> Length : 597
SequenceName : SEQ ID 418
25 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
30 <400> PreSequenceString :
atgaaaccaa atatgattgt aggagcatta gcgttaactt ctgtgtttat ggcaggtcac 60
ctacaggcgg ctgatggaac agtccatttc cgtggtgaaa ttattgacag tacttgcgaa 120
gtcactcctg aaactaaaga tcaggctcgtt gatttaggca aagtaaaccg tacagccttt 180
agtggcgctg atgatgtggc tgccccgacg gctttttcta tcgatctgac tcaatgcccg 240
35 gaaaccttta agtccgcgcg aattcgtttc gatggtaatg aagatgctca tggtaatggc 300
aacctggcaa ttggtacccc gctggataac tctaacgatg ctgccgctgg tattagcccg 360
agtataaca gtggggatta tactgggtgc ggtgccgtta gtgcagcgaa aggcgtagct 420
attcgtttat ataaccgtgc agataacact cagggtcaagt tatatgaaaa ttctgcatca 480
actccgattt ctaatggtaa tgcattcatg aagttcatgg ctcggtatat tgctacggaa 540
40 acgactattg accctgggtac agctaaccgc gactcgcagt ttacagttga atatataaaa 600
taa 603

<212> Type : DNA
<211> Length : 603
SequenceName : SEQ ID 419
45 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
50 <400> PreSequenceString :
gtgccaatTT tccagcgtga aggccatctc aaatatagct ttgccgcagg tgaatatcag 60
gccgggaatt atgacagcgc ctgcgcgcgt ttccgggcagc ttgatctgat ctacggttta 120
ccgtggggga tgacggccta cggcggcgta ttaatctcta ataattacaa tgcatttaca 180
ttagggatag ggaaaaactt tgggtatatc ggggcgattt ccattgatgt gacgcaggct 240
55 aaaagcgaac tgaataacga tcgcgatagc cagggacaat cttatcgttt cttatatattcc 300
aagagcttcg aaagcggcac cgatttcgcg cttgcgggct atcggtactc taccagcggg 360
ttctatacct tccaggaagc caccgatgtg cgcagtgcag ctgacagcga ctataaccgt 420
tatcacaagc gcagcgaaat acagggtaac ctgacgcagc aattaggggc ctatggctct 480
gtttatttaa atttaacgca gcaggattac tggaaacgac caggtaaaca gaacacggta 540
60 tcggcggggt acaacggacg tattggcaag gtcagttaca gtattgcata tagctggaat 600
aaaagccctg aatgggatga aagcgatcgc ttgtggtctt tcaatatattc cgttccacta 660
ggccggggct ggagtaacta tcgcgtcacg accgaccagg atggtcgtac caatcaacag 720
gttgggggtc gcggaacgct gcttgaggat cgcaacctga gctacagtgt ccaggaaggc 780
tacgccagca acggtgtggg taacagcggg aacgctaacg ttggctatca ggggtgggtcc 840
65 ggtaatgtca acgtaggcta tagctacggg aaagattacc ggcagctcaa ctacagcgtt 900
cgcgccggcg tgatagttca tagcgaaggc gtgacgcttt cccaaccgct aggcgaaacc 960
atgacgctca tctccgtacc cgggtgcgcgc aatgcccgcg tgggtgaataa cggcggcgctt 1020

cagggttgact ggatgggtaa cgcgatcgtg ccttatgcc a tgccgtatcg tgaaaacgaa 1080
atctcactgc gtagcgattc gttgggtgac gatgttgacg ttgaaaatgc gttccagaaa 1140
gtgggtgccaa cgcgtggagc gattgtcaga ggcggttttg ataccgcgt tggttaccgc 1200
gtattaatga cgtgcttcg ttccgcgggc agcccggtgc ctttggagc aacggcaacg 1260
5 ctaatcaccg ataaacaaaa cgaggtgagc agtatcgttg gtgaagaagg acagctctat 1320
attagcggaa tgccagagga aggacgggta ttgattaaat ggggtaata gaagtcgcag 1380
caatgcgtgg cgccttataa attatccctg gaattaaaac agggcggaat tattcctggt 1440
tcggccaatt gccagtaa 1458
<212> Type : DNA
10 <211> Length : 1458
SequenceName : SEQ ID 420
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgagtgggtt acaccgtcaa gcctcctacc ggagacagca atgagcagac acaattttatt 60
gattatattta atctgttcta cagtaagcgt gatcaggaac aaataagcat ctctcagcag 120
20 cttggaaatt acggtgcgac atttttcagt gccagtcgcc aaagtactg gaacacgtca 180
cgcagcgacc agcaaataatc atttggatta aatgtgccgt ttggtgatat tacgacttcg 240
ctgaattaca gctattccaa taatatatgg caaaacgatc gggatcattt actcgctttt 300
acgcttaatg ttcccttcag tcattggatg cgtacagaca gtcagtcggc atttcgtaat 360
tcaaacgcca gttacagtat gtcaaacgat ttgaaaggcg gcatgaccaa tctatcgggg 420
25 gtttatggca ctctgctgcc ggataataac ctgaattata gcgttcaggt cggtaacacc 480
cacggaggtg atacatcgtc tggcaccagt ggttacagta ctcttaatta tcgtggagct 540
tacggcaata ctaatgtcgg ttacagtcgg agtggtgaca gcagccagat ttattacgga 600
atgagtgggtg ggattattgc tcatgctgat ggcacacct ttggacagcc gctgggagac 660
acaatgggtc tgggttaaggc tcctggcgct gataatgtca aaatagagaa ccagaccgga 720
30 attcataccg actggcggtg ctatgccata ttaccatttg cgacagaata tagagaaaat 780
cgtgtcgcctc ttaacgcgaa ttcccttgca gataatgttg aactggatga aaccgtggtc 840
actgtcatcc caactcacgg tgctattgcc agagcaacat ttaatgcaca aatcggcggg 900
aaagtattaa tgacgttgaa gtacggtaat aaaagcgttc cattcggtgc aattgtcact 960
cacggagaga ataaaaatgg cagcattgtc gcggaaaacg gtcaggttta tctgactgga 1020
35 cttccacagt cagggaaatt acaggtttca tggggcaatg ataaaaactc aaactgtatt 1080
gtcgattaca agcttcctga agtctctcct ggaaccttgc tgaaccagca gacagcaatc 1140
tgtcgctaa 1149
<212> Type : DNA
<211> Length : 1149
40 SequenceName : SEQ ID 421
SequenceDescription :

Sequence
45 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgtctgctt tgtatgaacg ttcacagctg acgcaggtga tgatttcacg tgccccggcg 60
actgctgaaa ccatggagaa ggcggaatat ctgcgcctgg actgcacat caaggaagtc 120
cagttcaccg ccggtcagaa acaggatatt gatgtgacca cgtctgctc cacagagcag 180
50 gagaacatca acggtctggg ggcgtcgtcc gagatttcca tgcgggtaa tttttatctg 240
aatcaggccc agaacgcctt gcgtgatgcc tatgacaatg acacggtgta tgcgtttaag 300
gtgcagtttc cgtccggtaa gggctttaag ttccctggcg aagtgcgtca gcacacctgg 360
tcatccggta ccaacggcgt ggtggctgca acgttttcac ttgcctgaa gggtaaaccg 420
gtgtcctatg tggtaaccgt ggcgtttgtg aaaaatctgg ataagacact taccgtgaat 480
55 accggtgcgc tgctgacaat gtcagtcagt gtcaacgggg gaacgcggcc ttataaacac 540
gcctggaaga aggatgggtc gccggtagag ggacagacta ctgacacttt cagtaagcca 600
ggtgcgcagt caggtgataa gggggcttat acctgcgagg taacggattc tgcagaacag 660
ccgcagagca ttacctctga tgcgtgtaca gtaacgggtta atggtgcggg cggataa 717

60 <212> Type : DNA
<211> Length : 717
SequenceName : SEQ ID 422
SequenceDescription :

65 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :
atgagaaaca aaccttttta tcttctgtgc gcttttttgt ggctggcagt gagtcacgct 60
ttggctgcgg atagcacgat tactatccgc ggctatgtca gagataacgg ctgtagtgtg 120
gccgctgaat caaccaatth tactgttgat ctgatggaaa acgcggcgaa gcaatttaac 180
5 aacattggcg cgacgactcc tgcgtttcca tttcgtatth tgctgtcacc ctgtggtaac 240
gccgtttctg ccgtaaaagt tgggtttacc ggcgttgcag atagccacaa tgccaacctg 300
cttgcaactg aaaatacggg gtcagcggct tcgggactgg gaatacagct tctgaatgag 360
cagcaaaatc agataccctt taatgctcca tcgtccgcga tttcgtggac gaccctgacg 420
ccgggtaaac caaatacgtt gaatttttac gcccggttaa tggcgacaca ggtgcctgtc 480
10 actgcggggc atatcaatgc cacggctacc ttcactcttg aatatcagta a 531

<212> Type : DNA

<211> Length : 531

SequenceName : SEQ ID 423

15 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7

20 <400> PreSequenceString :
atgaataaat ccgttgtgtc aatttctgcg gcaatgttgg ttttactttg ccaaccgggc 60
atggggagcg aaatctcacc cgcaacaccg tcagatgaag acaactacac ctttgaccgc 120
caactcttcc gcggcagcag atttagtcag tcgtcattag caaaactgac aacacgtgag 180
tctgttgac cgggcaatta taaaatggat atctacacca acaataagtt gtcaggcagt 240
25 tggaatgtca cgttttaaga agccgctgat ggctcggttc tgccctgcct gacgcctgaa 300
gtcgcggacg cgataggcct caaaacaggg gaagataagg gggaaaaaga tcctgtctgt 360
acgtttgcta aggaactcgc tcccggcatc accagccaga cacagttgtc acaattgcgc 420
ctggacttat cggtgccaca gagtcaattg attagtcgcc ctgcgggcta tgttcccccc 480
agcgagctgg ataccggagc atcgctggcg ttcattgaatt atattgcca ctattacaac 540
30 gttgcctatt cagggcagaa tgctcatagc cagcggtcgc tatgggcac atttaattgg 600
ggcatcaacc ttgggtgcctg gcaatatcgt cagttatcca acatgacctg ggataatgac 660
aaagggaatc agtgaacaa tattcgtagc tatttgcaac gcccgctgcc cgccataaat 720
agccagttta tgatggggca gcttatcacc agcggaagat ttttctctgg actcagttat 780
cacggcggtta gtctcgcgac cgatgaacgt atgctgcgg actccatgcg cggctatgag 840
35 ccgactattc gcggcgtggc cgcaacaaac gccagagctc cggtaatgca aaacgggtcat 900
gaaatatatc agaccaccgt ggctcctggc ctttctgaga taaacgacct ataccacc 960
agctacagcg gcgatctgga tgtcaccgtt acggaagcta acggcgagc cagtcgtttc 1020
agtgtccctc tttcagccgt accagaatcg atgcgtccag gaacttccc ttataacgtg 1080
gaagtaggta aaacgcagga tagtggtgat gactcgatgt ttgggtgacct tacctggcag 1140
40 caggggatga ctaatacgtc gacatttaac agtgggtcgc gtatcgctga tggctaccag 1200
gcgctgatgc tgggcggagt ctatggcagt tcgctggggg catttggggc aaacctcact 1260
tggtcccatg cgcgtgttcc cgaaagcgaa gcgcagagt gttggatgtc gcaattaacc 1320
tggagtaaaa ctttccagcc tacttcaacc accgtctccc tggcaggtta tcgatactct 1380
accagcggct atcgtgatct ggctgatgtg ctgggagagc gtcattgctc cagcaataaa 1440
45 cagtcattgg actccagcca gtggcgtcaa cagtcgcgtc tcgactctac gtttaagtcag 1500
agccttgcca attacggcaa cctgtttgtg tcagggtcaa cacagaacta ccgtggcggc 1560
aagagccgtg atacacagct tcagttaggt tacagcaata gcttttagcca tggcatcagt 1620
atgaaccttt ccgtcggacg ccaaagaatg ggcggtata aagacaattc tgatgatatg 1680
cagacggtaa catccctttc attctcattc ccacttggcg gcaatggacc tcgtgtacca 1740
50 agtcttagca acagctggac ccattcaact gacggtagct cgcaattaca aagctcgcta 1800
accggaatgc ttgatgaagc acagaccacc aactacagcc tgaacgtcat gcgcgatcaa 1860
caatataagc agacgacgct tagcggaac atgcaaaaac gtttttcaca aactaccgtc 1920
ggattgaacg catcgaaggc ccaggattac tggcaggctt caggtaacgt acaaggcgcg 1980
atggctgtgc atgggtggcg cattactttc ggaccttacc tgggtgaaac gttcgccctg 2040
55 gtcgaagcta aaggcgagc aggtgcaaaa gtctataact ccagtcagct ggaaattaat 2100
gacagtggct atgcgcttgt tccggcagta acgcccatac gctacaaccg tatatctctc 2160
gatccacaag gaatggatgg cgatgccgag ttgggtcgaca gtgaaagaca ggtagcaccg 2220
gttgccgggt cggcggtgaa agtaattttc cgtaccgcgc ctggtaaaag gttgctgatt 2280
aaatcccgc tggcagatgg ttcggaactg ccaatgggag ccgatgtgct ggatgagaat 2340
60 aatacagtcg tcggtatagc cggtcagggg gggcaaatth acctccgcac agaacagaca 2400
aaaggccact tgtcagttcg ctgggggtgaa ggtgctaacg atagctgcca attgcccttt 2460
gatatcagcg ggaaggacag caatagccct atcatccgcc tgaatgaaac ctgtcagttc 2520
tga 2523

<212> Type : DNA

65 <211> Length : 2523

SequenceName : SEQ ID 424

SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7

5 <400> PreSequenceString :

```
atgaaactcg ccgctgttt tctgacactc cttcctggct tcgccgttgc cgccagctgg      60
acttctccgg ggttccctgc ctttagcgaa cagggaacgg gaacatttgt cagccacgcg      120
cagttgcca aaggtacgcg tccactcacg ctaaattttg accagcagtg ctggcagcct      180
gcagatgcga taaaactcaa tcagatgctt tccctgcaac cttgtagcaa cacgccgcct      240
10 caatggcgat tggttcaggga cggcaaatat acgctgcaaa tagaoacccg ctccggtacg      300
ccaacattga tgatttccat ccagaacgcc gccgaaccgg tagcaaacct ggtccgtgaa      360
tgcccgaat gggatggatt accgctcacg ctggatgtca gcgccacttt cccggaagga      420
gccgccgtac gggattatta cagccagcaa attgcgatag tgaagaacgg tcaaataacg      480
ttacaacccg ctgctaccag caacggttta. ctctgtctgg aacgggcaga aactgacgcc      540
15 tctgcccctt tcgactggca taacgccacg gtttactttg tgctgacaga tcgtttcgaa      600
aacggcgatc ccagtaatga ccagagttac ggacgtcata aagacgggat ggccgaaatt      660
ggcacttttc acggcggcga tttacggcgc ctgaccaaca aactggatta cctccagcag      720
ttgggcgtta atgctttatg gataagcgcc ccatttgagc aaattcacgg ctgggtcggc      780
ggcgggtacaa aaggcgattt cccgcattat gcctaccacg gttattacac acaggactgg      840
20 acgaatcttg atgccaatat gggcaacgaa gccgatctac ggacgctggg tgatagcgca      900
catcagcgcg gtattcgtat tctctttgat gtctgtatga accacaccgg ctatgccacg      960
ctggcggata tgcaggagta tcagtttggc gcgttatatc tttctgggtg cgaagtgaag      1020
aaaacgctgg gtgaacgctg gagcgactgg aaacctgccg ccgggcaaac ctggcatagc      1080
tttaacgatt acattaattt cagcgacaaa acaggctggg ataatgggtg gggaaaaaac      1140
25 tggatccgta ccgatatcgg cgattacgac aatcctggat tcgacgatct caccatgtcg      1200
ctagcctttt tgccggatat caaaaccgaa tcaactaccg cttctgggtc gccggtgttc      1260
tataaaaaaa aaacggatac ccacgctaaa gccatcgacg gctttacccc tcgcgattac      1320
ttaaccact ggttaagtca gtgggtccgc gactatggga ttgatgggtt tcgggtcgat      1380
accgcaaac atgttgagtt gccgccttgg cagcaactga aaaccgaagc cagcgccgcg      1440
30 cttcgcgaat ggaaaaaagc taaccccgac aaagcattag atgacaaacc tttctggatg      1500
accggtgaag cctggggcca cggcgtgatg caaagtgact actatcgcca cggcttcgat      1560
gcgatgatca atttcgatta tcaggagcag gcggcgaaag ctgtcgattg tattgcgcag      1620
atggatacga cctggcagca aatggcggag aaattgcagg gtttcaacgt gttgagctac      1680
ctctcgctgc atgatacccg tctgttccgt gaaggggcgc acaaagcagc agagttatta      1740
35 ctattagcgc caggcgcggt acaaactctt tatggcgatg aatcctcgcg tccgttcggg      1800
cctacaggtt ctgatccgct gcaaggatca cgttcggata tgaactggca ggatgttagc      1860
ggtaaatctg ccgccaacgt cgcgcactgg cagaaaatca gccagttccg cgcccgccat      1920
cccgcatttg gcgcgggcaa acaaacgaca ctttcgctga agcagggcta cggctttgtt      1980
cgtgagcatg gcgacgataa agtgctgggtc atctgggctg ggcaacagtg a          2031
```

<212> Type : DNA

<211> Length : 2031

SequenceName : SEQ ID 425

SequenceDescription :

45

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```
50 atgccacaac gacaccacca gggacataaa cgcacaccga aacagttggc gctcattatc      60
aaacgctggt tgccgatggg gctcactggc agcggcatgc tttgcactac cgctaacgcc      120
gaagagtatt atttcgaccc cattatgctg gaaaccacaa aaagtgggat gcaaacaacc      180
gatctgtcac gtttttcaaa aaaatacgca caactaccag gaacttatca ggttgatata      240
tggtctgaata aaaagaaggt ttcacagaaa aaaattacat ttaccgcaa tgcagagcaa      300
55 cttctgcagc cacagtttac ggtagaacaa ctacgtgagc tgggtattaa ggtggatgaa      360
atcccggcgc tggctgaaaa agatgacgat agcgtgatca actcgttga acaaatcatt      420
cccggtagac ctgctgaatt tgatttcaat catcagcgac ttaatttgag cattcccaa      480
attgcaactgt accgtgatgc aagaggttac gtctccctt ctcgttggga cgatggata      540
ccaacgctgt ttaccaacta ctcgtttaca ggttctgata accgttaccg ccagggcaat      600
60 cgtagccaac gacagtacct aaatatgcaa aatgggtgca attttggccc ctggcgatta      660
cgtaactatt ctacgtggac acgcaacgat caggcgtcaa gctggaacac tatcagtagt      720
tatttacaac gtgatataca ggcgttgaag tctcagttgc ttctgggaga aagcgccacc      780
agcggcagta ttttttccag ctacaacttt actggcgtgc aactcgcttc cgacgataat      840
atgttgccaa acagccagcg cggatttggc ccaacggtag gcggtatcgc aaacagtagt      900
65 gcaatcgtga ctatcaggca aaatgggtat gtgatctatc aaagcaacgt gccagcgggt      960
gcctttgaaa ttaacgatct ctaccctct tccaacagcg gcgatttaga agtcacgatt      1020
gaagaaagtg acggtacgca acgtcgcttt atccagcctt attcttcatt acccatgatg      1080
```


5 cagcgacctg ggcattctaaa atatagcgcg accgctggac gctatcgcg c gatgcaaac 1140
agtgatagca aggaacccga atttgctgaa gccacggcaa tatatggttt gaataatact 1200
tttacgctgt atggcgccct gctcggttct gaagattatt atgcgctggg gatcggtatc 1260
ggcggcacac ttggcgccact gggcgcggtt tcatggctat caatggcgta cgcagtacat caaagatatc 1320
ttcgataacc agcactcttt tcatggctat caatggcgta cgcagtacat caaagatatc 1380
ccggaaacca acaccaatat cgctgtcagc tactatcgct ataccaacga tggctatctt 1440
agttttgatg aagccaatac ccgcaattgg gactataaca gtcgccaaaa aagtgaattt 1500
caattcaaca tcagccagac aatatttgat ggggtaagtc tgtatgcctc cggttcacag 1560
caagactatt ggggcaataa cgagaaaaac aggaatatct ctggtggggg ttccggccag 1620
10 caatggggaa ttggttacag cctgaattat caatacagcc gctacactga tcaaaataat 1680
gaccgcgcac tctctttgaa tctcagtatt ccgttagaac gctgggtacc gcgtagccgg 1740
gtttcctatc agatgaccag ccagaaagat cgcccaaccc aacatgaaat gcgtcttgat 1800
ggctcactgc tggatgatgg tcgcctgagc tatagtctgg aacaaagtct ggatgacgat 1860
aacaaccata acagtagcgt gaacgccagt taccgttcac cttatggaac cttcagtgcg 1920
15 ggatacagtt acggtaatat cagtagccaa tacaattacg gcgttaccgg cggcgtgggt 1980
atccatcttc atggtgtgac gctctcgcaa tatctgggca acgcttttgc gcttattgat 2040
gctaaccggg catctggcgt gaggatacaa aactatccgg ggattgctac tgatcccttt 2100
ggctatgcag tggttcctta tctcacaact tatcaggaaa accgtctctc ggtagatact 2160
acgcagctgc ccgataacgt cgatcttgaa caaacaacac agtttgtggg gcccaacaga 2220
20 ggtgcaatgg tagcgcgcg tttcaacgcc aatatcggtt atcgcgact tgttacagtc 2280
agcgatcgca acggtaaacc gttgcccttt ggcgctcttg ccagcaacga tgatacgggg 2340
caacaaagta tcgtcgatga gggcggcata ctatatctct ctgggatata gagtaaatca 2400
caaagctgga ctgtacgctg gggaaatcag gcagatcaac aatgtcagtt tgcttttagt 2460
acaccggatt cagaaccaac aacctctgta ttacaaggca cagcgagtg ccattaa 2517

<212> Type : DNA

<211> Length : 2517

SequenceName : SEQ ID 426

SequenceDescription :

30

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

35 atgatgttca gaaatagaat attactaata tttatattgt gggctaattt tacctgggct 60
gggtgtcgta ctactgcac attaaatatt acagatggta ttaatgttgg ggagatttta 120
gcgaatgaaa cttccttttag taaaagtgtc gtgtttactg ggatatcttg tgatacgagc 180
acggataaaa tagtttataa aaatatccaa agtgattggg ttgaagtggg gccttttggg 240
aatggcgaaa aattaaagggt taaaatagag tctttaggta aaaccagcga cacaattggg 300
40 aaatccagca atgcgcaggc agtattacct tatgtgggta aaatagccag aggcacacct 360
gatttttactg gagaaagaaa atctacctgg tttatttcag ataccgtgat tgcaaataat 420
ggcgggtgagt catcgtcac catcgatttt tggttgggta tttgtaaggc attgaagttt 480
aactgggtgtg tgaattatct caccagcaaa ctggcggggg atacatttac gcttgggtta 540
aatatttcct attatcctaa aaatagcacc tgtaagcctg aaaacaccgt tataaaagta 600
45 gatgatatcg ccttggtcca gctcagaaat cagggaaaga ttgcggcgaa cagtaaggaa 660
ggaacaatta cgttgaaatg tgataatctt ttcggcgaca aaaaacaagc atcgcggaat 720
atggttgat atctttctag cagtgaacta gttaaaggaa gtaatactat tttgcgtggg 780
aaaacagata atggtgtagg gtttgtgttg gatctaacag aaccaccaa agggactgag 840
gctgccatta aaatttcggc caacggcgat cagggcgcg cgacatcatt atggaaaaca 900
50 gataaaccag gagtttcatt aaatagcaac attattaata taccagtcatt ggccagttac 960
tatgtatatg atgaaaaaaa agttaaatct ggcgactgg aagcaaccgc attaatcaac 1020
gtgaaatacg attaa 1035

<212> Type : DNA

<211> Length : 1035

55

SequenceName : SEQ ID 427

SequenceDescription :

Sequence

60 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

atgattaaaa aagcttcgct gctgacggcg tgtttctgtc cagccttttc cgcttgggca 60
caggatacca gcccggtac tctcgctggt actgctaacc gttttgaaca gcccgcgagc 120
actgtgcttg caccaaccac cggttgtagc cgtaaggata tcgaccgctg gcagtcgacc 180
65 tcggttaatg atgtgctgcg ccgtcttcog ggcgtcgata tcacccaaaa cggcggttca 240
ggtcagctct catctatctt tattcgcggt acaaatgcca gtcattgtgt ggtgttaatt 300
gatggcgtag gcctgaatct ggcggggggg agtgggtctg ccgaccttag ccagttccct 360

attgcgcttg tccagcgtgt tgaatatatc cgtggggccac gctccgcccgt ttatgggtcc 420
gatgcaatag gtgggggtggt gaatatcatc acgacgcgcg atgaaccggg aacggaaatt 480
tcagcaggggt ggggaagcaa tagttatcaa aactatgatg tctctacaca gcaacaactg 540
ggggataaga cacgagtaac gttgttgggc gattatgccc atactcatgg ttatgatgtt 600
5 gttgcctatg gtaataaccgg aacgcaagcg cagccagata acgatgggtt tttaagtaaa 660
acgctttatg gcgcgctgga gcataacttt actgatgcct ggagcggctt tgtgcgcggc 720
tatggctatg ataaccgtac caattatgac gcgtattatt ctccgggttc accattgggtc 780
gatacccgta aactctatag tcaaagttgg gacgccgggc tgcgatataa cggcgaactg 840
attaaatcac aactcattac cagctatagc catagcaaag attacaacta cgatccccat 900
10 tatggtcggt atgattcgtc ggcgacgctc gatgagatga agcaatacac cgtccagtgg 960
gcaaacaaca tcatcattgg ccacggtaat gttggtgcgg gtgttgactg gcagaagcag 1020
agcacggcac cgggcacagc ttatgttaag gatggatatg atcaacgtaa taccggcatc 1080
tatctgaccg ggctgcaaca agtcggcgat ttacctttg aaggcgcagc acgcagcgac 1140
gataactcac agtttggctc tcatggaacc tggcaaacca gcgcgggttg ggaattcatc 1200
15 gaaggttatc gcttcattgc ttctacggg acatcttata aggcaccaa tctggggcaa 1260
ctgtatggct tctacggaaa tccgaatctg gaccgggaga aaagcaaaca gtgggaaggc 1320
gcgtttgaag gcttaaccgc tggggtgaac tggcgtatct ccggatatcg taacgatgtc 1380
agtgacttga tcgattatga tgatcacacc ctgaaatatt acaacgaagg gaaagcgcgg 1440
attaagggcg tcgaggcgac cgccaatttt gataccggac cactgacgca tactgtgagt 1500
20 tatgattatg tcgatgcgcg caatgcaatt accgacacgc cgttggttac ccgtgctaaa 1560
cagcaggtga aataccagct cgactggcag ttgtatgact tcgactgggg tattacttat 1620
cagtatttag gcactcgcta tgataaggat tactcatctt atccttatca aaccgttaaa 1680
atgggcgggtg tgagcttgtg ggatcttgcg gttgcgtatc cggtcacctc tcacctgaca 1740
gttcgtggta aaatagccaa cctgttcgac aaagattatg agacagtcta tggctaccaa 1800
25 actgcaggac gggaatacac cttgtctggc agctacacct tctga 1845
<212> Type : DNA
<211> Length : 1845
SequenceName : SEQ ID 428
SequenceDescription :
30
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 atgaaaaaca aattgttatt tatgatgtta acaatactgg gtgcgcctgg gattgcagcc 60
gcagcaggtt atgatttagc taattcagaa tataacttcg cggtaaatga attgagtaag 120
tcttcattta atcaggcagc cataattggg caagctggga ctaataatag tgctcagtta 180
cggcagggag gctcaaaact tttggcgggt gttgcgcaag aaggtagtag caaccgggca 240
aagattgacc agacaggaga ttataacctt gcataattg atcaggcggg cagtgccaat 300
40 gatgccagta ttctgcaagg tgcttatggg aatactgcga tgattatcca gaaagggtct 360
ggtaataaag caaatattac acagtatggg actcaaaaaa cggcaattgt agtgcagaga 420
cagtcgcaaa tggctattcg cgtgacacaa cgtaa 456
<212> Type : DNA
<211> Length : 456
45 SequenceName : SEQ ID 429
SequenceDescription :
Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaacattt ttgcatattt actggtactt gtattttcca tgagcatgag cagcagcgcg 60
tttgccagcg tggtaatgac eggaaccggt attattttcc ctggtgacgc aaaggaaaaa 120
accatccagt tgcgaaatac cagcgatcag ccctatatca ttaatatcca tgttgaggat 180
55 gaacgtggtt ctgacaagaa tgtaccgttt atgccaaccc cgcagacatt tcgcatggaa 240
gctgccgcag gtcaggcggt acgcctgctc tacactggta ataatttacc gcaggatcgc 300
gagtcgtgtt tctggtttag tttcagtcaa ctaccttacc tgaataagaa tgataaaagt 360
cagaaccagc tcatcctggc cctgactaat cgagtcaaaa ttttctatcg tcccagctcg 420
attgtcggta aatccagtga cgcacccaaa aacctgactt accaggtaaa acagaaccgc 480
60 attgaagtga cgaatccac gggctattac gtcacaattc gcgccgctga actgcttaat 540
aatggtaaaa aagtccccct cgcgaattcg gtaatgattg ctctcctcaa cacaactgaa 600
tggaactac cctctggcat cagtgtcgct cccggcgcc agatccattt agtgaccgtc 660
aacgactatg gcgtaaatgt tacgtctgag catgccttat aa 702
<212> Type : DNA
65 <211> Length : 702
SequenceName : SEQ ID 430
SequenceDescription :